

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 618 353**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.09.2008 PCT/US2008/075528**

87 Fecha y número de publicación internacional: **03.09.2009 WO09108215**

96 Fecha de presentación y número de la solicitud europea: **06.09.2008 E 08872935 (5)**

97 Fecha y número de publicación de la concesión europea: **14.12.2016 EP 2195467**

54 Título: **Clasificación de tumores y pronóstico del cáncer de mama**

30 Prioridad:

06.09.2007 US 970529 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

21.06.2017

73 Titular/es:

**BIOETHERANOSTICS, INC. (100.0%)
9640 Towne Centre Drive, Suite 200
San Diego, CA 92121, US**

72 Inventor/es:

**ERLANDER, MARK, G.;
MA, XIAO-JUN y
SGROI, DENNIS**

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 618 353 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Clasificación de tumores y pronóstico del cáncer de mama

5 **Campo de la invención**

La invención se refiere a la identificación y al uso de perfiles de expresión génica, o patrones, con relevancia clínica para el cáncer. En particular, la invención se basa en parte en las identidades de los genes que se expresan en correlación con el grado tumoral. Los niveles de expresión génica forman un índice molecular que es capaz de determinar el grado tumoral en un paciente y predecir el resultado clínico y, por tanto, el pronóstico para el paciente. La clasificación molecular del cáncer puede utilizarse opcionalmente en combinación con un segundo índice molecular para el diagnóstico del cáncer y su pronóstico.

Los perfiles de expresión génica, ya sea que plasmados en la expresión de ácidos nucleicos, la expresión de proteínas u otros formatos expresión, pueden utilizarse para predecir el resultado clínico de los sujetos que padecen cáncer, predecir la reaparición del cáncer y/o predecir la aparición de cáncer metastásico. Los perfiles también pueden utilizarse en el estudio y/o el diagnóstico de las células y el tejido cancerosos, así como para el estudio del pronóstico de un sujeto. Cuando se utilizan para el diagnóstico o el pronóstico, los perfiles se utilizan para determinar el tratamiento del cáncer basado en la probabilidad de esperanza de vida, de reaparición del cáncer y/o de metástasis del cáncer.

Antecedentes de la invención

Los estudios de creación perfiles de expresión del genoma completo han creado una "pequeña inundación" de firmas genéticas de pronóstico para el cáncer de mama. Una cuestión importante es si estas firmas se solapan en el espacio de pronóstico y si la combinación de varias de ellas proporcionaría un pronóstico más preciso. En un estudio comparativo, se descubrió que cuatro firmas (los subtipos intrínsecos, la firma de 70 genes, la firma de respuesta a la herida y el índice de recurrencia), desarrolladas utilizando diferentes cohortes de pacientes y metodologías, eran muy concordantes en la clasificación de los pacientes en grupos de bajo y alto riesgo. Además, la combinación de estas firmas no proporciona una mejora significativa en la precisión predictiva, lo que señala que el espacio de información de pronóstico abarcado por estas firmas se solapa en gran parte.

La importancia en el pronóstico del grado tumoral también ha sido establecido (Cianfrocca M, Goldstein LJ *Oncologist* 9:606-16, 2004). Diversos índices moleculares para el pronóstico del cáncer han sido citados anteriormente. Los ejemplos incluyen un índice de grado genómico (IGG) basado en 97 genes asociados al grado tumoral, que se ha demostrado que es fuertemente útil como pronóstico (Sotiriou C, Wirapati P, Loi S, *et al. J Natl Cancer Inst* 98:262-72, 2006); una firma de 70 genes (van 't Veer LJ, Dai H, van de Vijver MJ, *et al. Nature* 415:530-6, 2002); y el algoritmo del índice de recurrencia de 21 genes Oncotype DX (Paik S, Shak S, Tang G, *et al. N Engl J Med* 351:2817-26, 2004).

Se ha citado que la firma del grado tumoral de 97 genes es comparable a la firma de 70 genes y al algoritmo del índice de recurrencia en cohortes independientes y se ha planteado la hipótesis de que la mayor parte del poder de pronóstico de estas firmas proviene de genes asociados a la proliferación celular.

Una comparación de las firmas descritos anteriormente señala que los genes relacionados con el grado tumoral son denominadores comunes de estas firmas (véase, por ejemplo, Sotiriou C, Wirapati P, Loi S, *et al. J Natl Cancer Inst* 98:262-72, 2006; Desmedt C, C Sotiriou *Cell Cycle* 5:2198-202, 2006; Loi S, Haibe-Kains B, Desmedt C, *et al. J Clin Oncol* 25:1239-46, 2007; y Sotiriou C, Piccart MJ *Nat Rev Cancer* 7:545-53, 2007).

Más recientemente, se ha propuesto que una "firma génica de invasividad" (FGI) de 186 genes, derivado de la comparación de las células de cáncer de mama CD44+CD24-/bajo oncógenas con el epitelio mamario normal, se extiende más allá del espacio de pronóstico basado en la proliferación. Sin embargo, un examen cuidadoso señala que su capacidad de pronóstico también puede derivar de genes relacionados con la proliferación ya que el DGI está altamente correlacionado con una firma de grado tumoral ($r = 0,81$).

Dada la importancia del grado tumoral en el pronóstico y la existencia de cientos de genes cuyos niveles de expresión se correlacionan altamente con el grado y la proliferación tumorales, puede que no sea sorprendente que puedan desarrollarse una multitud de firmas pronósticas aparentemente distintas. Además, la robustez del pronóstico y la redundancia de estos genes señalan que un ensayo mucho más simple que implique unos pocos genes puede ser suficiente. Por ejemplo, se ha observado que solo se necesita una fracción de los 97 genes para el IGG para el pronóstico. En un estudio independiente, Ivshina *et al.* también demostró que una firma de grado tumoral de 264 genes puede reducirse a 6 genes con medios informáticos (Ivshina AV, George J, Senko S, *et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res* 66:10292-301, 2006).

La cita de documentos en el presente documento no ha de interpretarse como que refleja una admisión de que alguno es estado del arte relevante. Además, su cita no es una indicación de una búsqueda de divulgaciones relevantes. Todas las declaraciones relativas a la fecha o fechas o al contenido de los documentos se basan en la información disponible y no son una admisión en cuanto a su exactitud o corrección.

5 Yamamoto *et al*, *Cancer Genetics and Cytogenetics*, Vol 174, 2007, 42-47 describe que la sobreexpresión de BUBR1 predijo la reaparición del tumor y la progresión de la enfermedad en el cáncer de vejiga.

10 El documento WO 2005/008213A describe un algoritmo de perfil de expresión y un ensayo para el pronóstico del cáncer.

Breve resumen de la invención

15 La invención se basa en parte en el descubrimiento y la determinación de los niveles de expresión génica en las células tumorales que se correlacionan con el grado tumoral. Además del uso de los niveles de expresión de los genes identificados como una firma del grado tumoral, los niveles de expresión pueden utilizarse para proporcionar información de pronóstico, tal como la reaparición del cáncer, e información predictiva, tal como el grado de respuesta a ciertas terapias.

20 Un gen identificado por la invención codifica Bub1B. Por tanto, se describen composiciones y métodos para el uso de la expresión del gen Bub1B para estudiar o determinar el grado tumoral, para proporcionar información de pronóstico y/o proporcionar predicciones del grado de respuesta clínica. En algunos casos, la determinación se hace con las células tumorales de un sujeto para llegar a un diagnóstico correspondiente al de un tumor de grado I, de grado III o grado intermedio. Los ejemplos no limitantes de células para su uso en la invención incluyen las recién aisladas del sujeto, las congeladas después de su aislamiento y las que se fijan y/o se incluyen, tales como las fijadas con formol, incluidas en parafina (FFIP). En algunas realizaciones, las células son células de la mama, tales como células de cáncer de mama.

30 En un segundo aspecto, se desvelan composiciones y métodos para el uso niveles de expresión de otros cuatro genes para determinar el grado tumoral, para proporcionar información de pronóstico y/o proporcionar predicciones del grado de respuesta clínica. Estos genes adicionales codifican CENPA, NEK2, RACGAP1 y RRM2. Por tanto, la invención se basa en parte en el descubrimiento de cinco genes, cuyos niveles de expresión son útiles para la determinación del grado tumoral en un sujeto que padece cáncer y para proporcionar determinaciones de pronóstico y predictivas para el sujeto.

35 Mientras que el nivel de expresión de cada uno de estos cinco genes puede utilizarse por separado en el estudio o la determinación del grado tumoral o para proporcionar información adicional, un tercer aspecto de la invención incluye el uso de cualquier combinación de los cinco genes divulgados. De este modo, en algunas realizaciones, puede utilizarse una combinación de niveles de expresión de Bub1B y uno, dos o tres cualesquiera de estos cuatro genes adicionales. De forma similar, puede utilizarse una combinación de niveles de expresión de CENPA y uno, dos o tres cualesquiera de Bub1B, NEK2, RACGAP1 o RRM2; de NEK2 y uno, dos o tres cualesquiera de Bub1B, CENPA, RACGAP1 o RRM2; de RACGAP1 y uno, dos o tres cualesquiera de Bub1B, CENPA, NEK2 o RRM2; de RRM2 y uno, dos o tres cualesquiera de Bub1B, CENPA, NEK2 o RACGAP1.

45 En una realización, se desvela una combinación de los cinco niveles de expresión, como una firma de grado tumoral de 5 genes (o índice de grado molecular). Este índice, o IGM, es capaz de recapitular el grado tumoral y predecir el resultado clínico con un rendimiento comparable al IGG de 97 genes en dos cohortes independientes. El IGM también sirve como un factor de pronóstico de reaparición del cáncer y/o del resultado de supervivencia.

50 En un aspecto adicional, la invención incluye el uso del IGM de 5 genes en combinación con un segundo índice molecular para el cáncer. En una realización, la combinación es del segundo índice molecular con los cinco genes divulgados. En otras realizaciones, la combinación puede ser con uno, dos, tres o cuatro de los cinco genes divulgados como se describe en el presente documento. En algunos casos, el segundo índice molecular es uno basado en los niveles de expresión de dos genes HoxB 13 e IL17BR. En particular, puede utilizarse un ratio de dos genes de la expresión de HoxB13 frente a la expresión de IL17BR (o ratio HoxB13:IL17BR) como segundo índice molecular (véase el documento US 2005/0239079 A1, el documento US 2005/0239083 A1 y el documento US 2006/0154267 A1). En una realización alternativa, el segundo índice puede ser un ratio de dos genes de la expresión de HoxB13 frente a la expresión de CHDH.

60 El ratio HoxB13:IL17BR (H:I) se descubrió sobre la base de un estudio de nuevos biomarcadores predictivos del resultado clínico más allá de los factores de pronóstico convencionales. Se emparejaron pacientes que desarrollaron reapariciones de cáncer con aquellos que no las desarrollaron con respecto al estadio y el grado tumoral. Se descubrió el ratio simple H:I era adecuada para predecir la reaparición del cáncer en pacientes con cáncer de mama con receptor estrogénico positivo (RE+) que recibían terapia adyuvante con tamoxifeno. Los estudios posteriores (Ma XJ, Hilsenbeck SG, Wang W, et al *J Clin Oncol* 24:4611-9, 2006; Goetz MP, Suman VJ, Ingle JN, et al *Clin Cancer Res* 12:2080-7, 2006; Jerevall PL, Brommesson S, Strand C, et al *Breast Cancer Res Treat*, 2007; y Jansen

MP, Sieuwerts AM, Look MP, et al *J Clin Oncol* 25:662-8, 2007) han demostrado, adicionalmente, que el ratio es tanto de pronóstico, tal como por ser un indicador de la agresividad tumoral, como predictiva del beneficio del tamoxifeno (es decir, la respuesta/resistencia al tamoxifeno) en ensayos clínicos tanto retrospectivos como

5 Cuando tanto el IGM de 5 genes divulgados como el ratio H:I se analizaron utilizando la transcripción inversa-reacción en cadena de la polimerasa (RT-PCR) en tiempo real, se descubrió que la combinación proporciona una estratificación superior del riesgo de reaparición que la que es posible mediante cualquiera de ellos solo. Esto refleja un descubrimiento inesperado porque indica que el ratio H:I es independiente del grado tumoral. Como tal, la
10 combinación de los dos índices mejora el diagnóstico del cáncer y permite una determinación más precisa de su pronóstico mediante el análisis de forma eficiente de parámetros independientes pertinentes para el cáncer.

En realizaciones alternativas, la expresión del uno o más de los genes en la firma de 5 genes divulgada puede utilizarse en combinación con otros genes u otro índice molecular para el pronóstico del cáncer. Los ejemplos no
15 limitantes incluyen el índice de grado genómico (IGG) basado en 97 genes asociados al grado tumoral (Sotiriou *et al*, citado anteriormente) y los subconjuntos de genes dentro de esos 97 genes; la firma de 70 genes MammaPrint (van 't Veer *et al*, citado anteriormente) y los subconjuntos de genes dentro de esos 70 genes; el algoritmo del índice de recurrencia de 21 genes OncotypeDX (Paik *et al*, citado anteriormente) y los subconjuntos de genes dentro de esos 21 genes; y el ensayo de 76 genes Veridex (Wang *et al Lancet*, 365(9460):671-679, 2005) y los subconjuntos de
20 genes dentro de esos 76 genes. En otros casos, la expresión de uno o más de la firma de grado tumoral de 5 genes divulgada puede utilizarse en combinación con el nivel o niveles de expresión de uno o más genes expresados en correlación con un fenotipo de proliferación. En algunos casos, los genes expresados en correlación con un fenotipo de proliferación están dentro de los conjuntos de 97, 70, 21 y 76 genes como se han descrito anteriormente. Son ejemplos no limitantes de los genes expresados en correlación con un fenotipo de proliferación Ki-67, STK15, Survivina, Ciclina B1 y MYBL2. De este modo el nivel o niveles de expresión de uno, dos, tres o cuatro cualesquiera de los cinco genes del IGM pueden utilizarse con otros genes u otro índice molecular para el pronóstico del cáncer o como un factor predictivo del resultado clínico. Por supuesto, los niveles de expresión de los cinco genes, como un IGM, también pueden utilizarse en combinación con genes adicionales u otro índice como se ha descrito anteriormente y a continuación en el presente documento. Adicionalmente, una combinación de niveles de expresión
25 de uno, algunos o todos los genes del IGM con genes adicionales también puede combinarse adicionalmente con el ratio H:I como se ha descrito anteriormente y a continuación en el presente documento como un factor de pronóstico o un factor predictivo del resultado clínico.

De este modo, las realizaciones de la invención incluyen métodos que someten a ensayo la expresión de uno, algunos o todos los genes del IGM, opcionalmente con uno o más genes adicionales como se ha descrito
35 anteriormente y opcionalmente en combinación con el ratio H:I, como un factor de pronóstico o un factor predictivo del resultado del tratamiento. Un método de ensayo de este tipo puede utilizarse para estratificar los sujetos RE+ por su valor de pronóstico y su valor predictivo. Como pronóstico, la estratificación puede basarse en los niveles de expresión diferenciales que se correlacionan con, y de este modo indican, la agresividad tumoral como ejemplo no limitante. Como factor predictivo, la estratificación puede basarse en los niveles de expresión diferenciales que se correlacionan con, y de este modo indican, el grado de respuesta (o sensibilidad) y/o la falta de respuesta (o resistencia) a la quimioterapia, que también puede considerarse como un factor predictivo del beneficio de la quimioterapia. Como ejemplo no limitante, la estratificación (basada en los niveles de expresión) puede utilizarse para predecir la resistencia endocrina (tal como la resistencia al tamoxifeno como ejemplo no limitante) y/o la predicción de los beneficios de los inhibidores que se dirigen a cánceres de mama endocrino resistentes. Los ejemplos no limitantes de dichos inhibidores incluyen los que se dirigen a mTOR (diana en mamíferos de la rapamicina, una serina/treonina proteína cinasa), PI3K (fosfoinosítido 3-cinasa), una serina/treonina proteína cinasa de la familia AKT (cuyos miembros incluyen Akt1, Akt2 y Akt3 en los seres humanos) y/o EGFR (receptor del factor de crecimiento epidérmico; HER1 en los seres humanos). La detección de la expresión génica puede ser, por
45 supuesto, en cualquier muestra adecuada que contenga células como se describe en el presente documento.

En realizaciones alternativas adicionales de la divulgación, el ratio H:I independiente del grado tumoral puede utilizarse en combinación con un índice molecular diferente para el pronóstico del cáncer (en lugar de uno, algunos o todos los genes del IGM). Los ejemplos no limitantes de dichos índices incluyen el índice de grado genómico (IGG) basado en 97 genes asociados al grado tumoral (Sotiriou *et al*, citado anteriormente) y los subconjuntos de genes dentro de esos 97 genes; la firma de 70 genes MammaPrint (van 't Veer *et al*, citado anteriormente) y los subconjuntos de genes dentro de esos 70 genes; el algoritmo del índice de recurrencia de 21 genes OncotypeDX (Paik *et al*, citado anteriormente) y los subconjuntos de genes dentro de esos 21 genes; y el ensayo de 76 genes Veridex (Wang *et al Lancet*, 365(9460):671-679, 2005) y los subconjuntos de genes dentro de esos 76 genes. En
60 otros casos, puede utilizarse el ratio H:I en combinación con el nivel o niveles de expresión de uno o más genes expresados en correlación con un fenotipo de proliferación. En algunos casos, los genes expresados en correlación con un fenotipo de proliferación están dentro de los conjuntos de 97, 70, 21 y 76 genes como se han descrito anteriormente. Son ejemplos no limitantes de los genes expresados en correlación con un fenotipo de proliferación Ki-67, STK15, Survivina, Ciclina B1 y MYBL2.

65 De este modo, en algunas realizaciones, la invención incluye someter a ensayo la expresión del ratio H:I en

combinación con el nivel o niveles de expresión de uno o más genes, tales como uno o más seleccionados entre Ki-67, STK15, Survivina, Ciclina B1 y MYBL2. El método de ensayo puede utilizarse para estratificar los sujetos RE+ por su valor de pronóstico y su valor predictivo. Como pronóstico, la estratificación puede basarse en los niveles de expresión diferenciales que se correlacionan con, y de este modo indican, la agresividad tumoral como ejemplo no limitante. Como factor predictivo, la estratificación puede basarse en los niveles de expresión diferenciales que se correlacionan con, y de este modo indican, el grado de respuesta (o sensibilidad) y/o la falta de respuesta (o resistencia) a la quimioterapia, que también puede considerarse como un factor predictivo del beneficio de la quimioterapia. Como ejemplo no limitante, la estratificación (basada en los niveles de expresión) puede utilizarse para predecir la resistencia endocrina y/o la predicción de los beneficios de los inhibidores que se dirigen a cánceres de mama endocrinos resistentes. Los ejemplos no limitantes de dichos inhibidores incluyen los que se dirigen a mTOR (diana en mamíferos de la rapamicina, una serina/treonina proteína cinasa), PI3K (fosfoinosítido 3-cinasa), una serina/treonina proteína cinasa de la familia AKT (cuyos miembros incluyen Akt1, Akt2 y Akt3 en los seres humanos) y/o EGFR (receptor del factor de crecimiento epidérmico; HER1 en los seres humanos). La detección de la expresión génica puede ser, por supuesto, en cualquier muestra adecuada que contenga células como se describe en el presente documento.

En un aspecto adicional, la expresión de uno o más genes seleccionados entre Bub1B, CENPA, NEK2, RACGAP1 y RRM2 puede utilizarse como un factor de pronóstico o un factor predictivo del resultado clínico o para determinar el grado tumoral en un sujeto con una enfermedad mamaria benigna, tal como un sujeto al que se le diagnosticaría una enfermedad mamaria benigna en ausencia de la presente invención. El papel importante de la enfermedad mamaria benigna se analiza por Hartmann *et al.* (*N. Engl. J. Med.* 2005, 353:3). Los ejemplos no limitantes de enfermedad mamaria benigna incluyen los hallazgos histológicos de lesiones no proliferativas, lesiones proliferativas sin atipia e hiperplasia atípica.

Dada la observación de que el cáncer de mama se produce después de un diagnóstico de enfermedad mamaria benigna, se ha especulado que los precursores de cáncer de mama están presentes en algunos casos de enfermedad mamaria benigna, tales como los que implican lesiones con atipia o hiperplasia atípica. De este modo, la presente invención incluye un método para determinar el grado tumoral en una célula mamaria de un sujeto, tal como una célula de una muestra histológica utilizada para diagnosticar una enfermedad mamaria benigna. El método puede comprender someter a ensayo una muestra de células mamarias de un sujeto para determinar los niveles de expresión de Bub1B, CENPA, NEK2, RACGAP1 y RRM2, en el que dichos niveles de expresión se correlacionan con un tumor de grado I o grado III o incluso un tumor de grado intermedio. Como alternativa, el método puede comprender someter a ensayo cualquier subconjunto de estos cinco genes, hasta solo uno cualquiera de los genes, para determinar la posible presencia de células tumorales de grado I, de grado III o de un grado intermedio. En algunas realizaciones, las células son de la muestra utilizada para diagnosticar la presencia de lesiones con atipia o hiperplasia atípica.

Por supuesto, la invención incluye adicionalmente el uso del IGM y del ratio H:I en una muestra de un sujeto con una enfermedad mamaria benigna para determinar si el sujeto está en riesgo de desarrollo posterior de cáncer de mama. Como alternativa, la invención prevé el uso de solamente el ratio H:I con una muestra de este tipo para determinar el riesgo de desarrollo de cáncer de mama.

En un aspecto adicional, puede utilizarse la expresión de uno o más genes seleccionados entre Bub1B, CENPA, NEK2, RACGAP1 y RRM2 como un factor de pronóstico o un factor predictivo del resultado clínico o para determinar el grado tumoral en un sujeto con carcinoma ductal *in situ* (o CDIS). Por tanto, la presente invención incluye un método para determinar el grado tumoral en una célula mamaria de un sujeto que padece, o se sospecha que tiene, CDIS. La célula puede ser una de una muestra histológica utilizada para diagnosticar el CDIS en el sujeto. El método puede comprender someter a ensayo una muestra de células de cáncer de mama de un sujeto para determinar los niveles de expresión de Bub1B, CENPA, NEK2, RACGAP1 y RRM2, en el que dichos niveles de expresión se utilizan como un factor de pronóstico o un factor predictivo del resultado clínico o se correlacionan con un tumor de grado I o de grado III o incluso un tumor de grado intermedio. Como alternativa, el método puede comprender someter a ensayo cualquier subconjunto de estos cinco genes, hasta solo uno cualquiera de los genes, como un factor de pronóstico o un factor predictivo del resultado clínico o para determinar la posible presencia de células tumorales de grado I, de grado III o de un grado intermedio.

En otro aspecto, la invención incluye composiciones y métodos para detectar la expresión de uno o más genes seleccionados entre Bub1B, CENPA, NEK2, RACGAP1 y RRM2 para su uso como un pronóstico de la reaparición local de cáncer en el CDIS. En algunas realizaciones, un método basado en los niveles de expresión se utiliza ventajosamente en una muestra que contiene células de cáncer de mama de un sujeto con CDIS. Como ejemplo no limitante, la célula puede ser una de una muestra histológica preoperatoria utilizada para diagnosticar el cáncer en el sujeto. Para un sujeto de este tipo, el tratamiento de referencia es la cirugía, prefiriéndose la cirugía conservadora de la mama a una mastectomía radical, para retirar el CDIS. Esto con frecuencia va seguido de radioterapia postoperatoria, opcionalmente con terapia endocrina, tal como el tratamiento con tamoxifeno, un modulador selectivo del receptor estrogénico (SERM, por sus siglas en inglés), un regulador negativo selectivo del receptor estrogénico (SERD, por sus siglas en inglés) o un inhibidor de la aromatasa (AI, por sus siglas en inglés) y/o con quimioterapia. Pero este protocolo para abordar la posibilidad de reaparición del cáncer conduce a un exceso de tratamiento en

muchos sujetos que no van a experimentar reaparición del cáncer y al fracaso, en los casos de reaparición del cáncer.

5 Por tanto, la invención incluye la detección de la expresión de los cinco de estos genes donde la alta expresión del IGM es un indicador del aumento de la probabilidad de reaparición local del cáncer en el sujeto debido a un fallo de la cirugía conservadora de la mama y la radioterapia y/o terapia endocrina o quimioterapia posteriores. En otras realizaciones, un método de este tipo utiliza la detección del ratio H:I como indicador sustituto de la alta expresión del IGM. Por supuesto, la invención incluye un método que combina la detección del IGM alto y el ratio H:I como indicadores del aumento de la probabilidad de reaparición local del cáncer después del tratamiento para el CDIS. 10 Como alternativa, el método puede comprender someter a ensayo cualquier subconjunto de estos cinco genes del IGM, hasta solo uno cualquiera de los genes, opcionalmente en combinación con el ratio H:I, como indicadores del aumento de la probabilidad de reaparición local del cáncer después del tratamiento para el CDIS.

15 Los métodos pueden incluir adicionalmente la identificación del sujeto como que tiene probabilidades, o no tiene probabilidades, de experimentar la reaparición local del cáncer y, opcionalmente, pueden incluir adicionalmente el ajuste de las modalidades de tratamiento para el sujeto para abordar el resultado esperado. Como ejemplo no limitante, la determinación de una baja probabilidad de reaparición puede utilizarse para confirmar la idoneidad de, o para seleccionar, la cirugía conservadora de la mama, opcionalmente con una reducción de las terapias postoperatorias, tal como la omisión de radiación y/o la omisión de tratamiento endocrino o quimioterapia. Como otro ejemplo no limitante, la determinación de una alta probabilidad de reaparición puede utilizarse para confirmar la idoneidad de, o para seleccionar, la mastectomía radical con inclusión de terapias postoperatorias, tales como la radiación y/o la terapia endocrina o la quimioterapia. 20

25 En un aspecto adicional más, la invención incluye el uso de uno o más genes seleccionados entre Bub1B, CENPA, NEK2, RACGAP1 y RRM2 (o los cinco genes) como un factor de pronóstico o un factor predictivo del resultado clínico o para determinar el grado tumoral en un sujeto que está en evaluación basado en el consenso de expertos de St. Gallen de 2005 sobre el tratamiento primario del cáncer de mama temprano (véase Goldhirsch *et al. Ann. Oncol.* 2005, 16:1569-1583). Por tanto, la presente invención incluye un método para evaluar la expresión de uno, algunos o todos los genes del IGM en una célula mamaria de un sujeto como parte del diagnóstico diferencial y la selección de la terapia basada en el consenso de expertos. Los ejemplos no limitantes de partes del consenso que pueden utilizarse con los métodos divulgados incluyen el algoritmo para la selección de la terapia sistémica adyuvante para el cáncer de mama temprano; el grado de respuesta o la falta de respuesta a la terapia endocrina o el grado incierto de respuesta endocrina; y el estado ganglionar. Por supuesto, también se prevé la inclusión de uno o más aspectos de la invención en el consenso como un todo. En otras realizaciones, los métodos divulgados de creación de perfiles de expresión génica molecular se utilizan para confirmar las clasificaciones de grupos de bajo y alto riesgo, así como resolver incluir al menos algunos sujetos de la categoría de riesgo intermedio en los grupos de riesgo bajo o alto. 30 35

40 En algunos casos, los métodos divulgados pueden utilizarse para seleccionar o eliminar terapias para las mujeres premenopáusicas o para las mujeres posmenopáusicas, con diagnóstico de cáncer. Las mujeres premenopáusicas incluyen aquellas que tienen menos de aproximadamente 35 años de edad. En estos sujetos, la alta expresión del IGM es un indicador de reaparición del cáncer. De este modo, la invención incluye el uso del nivel o niveles de expresión de uno o más genes seleccionados entre Bub1B, CENPA, NEK2, RACGAP1 y RRM2 como pronóstico de reaparición del cáncer de mama, tal como en los casos de CDIS, en un sujeto premenopáusico. Opcionalmente, el ratio H:I también se somete a ensayo y se utiliza como una combinación con el gen o genes del IGM. El método puede incluir someter a ensayo una muestra que contiene célula de cáncer de mama de un sujeto para determinar la expresión de estos genes. Como ejemplo no limitante, la célula puede ser una de una muestra histológica preoperatoria utilizada para diagnosticar el cáncer en el sujeto. En otros casos, el método incluye el uso de la expresión de los cinco de estos genes como una realización, donde la alta expresión del IGM es un indicador del aumento de la probabilidad de reaparición del cáncer en el sujeto premenopáusico. 45 50

55 Los métodos pueden incluir la identificación del sujeto premenopáusico como que tiene probabilidades, o no tiene probabilidades, de experimentar la reaparición local del cáncer y, opcionalmente, pueden incluir adicionalmente el ajuste de las modalidades de tratamiento para el sujeto para abordar el resultado esperado. Como ejemplo no limitante, la determinación de una baja probabilidad de reaparición puede utilizarse para confirmar la idoneidad de, o para seleccionar, la cirugía conservadora de la mama, opcionalmente con una reducción de las terapias postoperatorias, tal como la omisión de radiación y/o la omisión de tratamiento endocrino o quimioterapia. Como otro ejemplo no limitante, la determinación de una alta probabilidad de reaparición puede utilizarse para confirmar la idoneidad de, o para seleccionar, la mastectomía radical con inclusión de terapias postoperatorias, tales como la radiación y/o la terapia endocrina o la quimioterapia. 60

65 En otros casos, los métodos pueden utilizarse para ayudar en la selección del tratamiento, tal como entre la terapia endocrina, la quimioterapia, la radioterapia o una combinación de las mismas. En algunas realizaciones, la invención incluye composiciones y métodos para determinar los niveles de expresión de uno o más de los cinco genes del IGM, o de los cinco, como un factor predictivo de la eficacia de la terapia endocrina. En algunos casos, el factor predictivo puede ser del grado de respuesta o de la falta de respuesta a un SERM, tal como tamoxifeno, o a un

SERD. Esto incluye los casos en los que el ensayo de una muestra que contiene células de cáncer de mama de un sujeto revela un alto IGM, lo que indica la probabilidad de falta de respuesta al tamoxifeno. En otros casos, el factor predictivo puede ser de la eficacia de una forma de terapia endocrina sobre otra. Esto incluye un método que determina los niveles de expresión de uno, algunos o todos los genes del IGM como un indicador de un mayor grado de respuesta a un inhibidor de la aromatasas (AI) en comparación con el tamoxifeno u otro SERM o un SERD. El método puede incluir la identificación de un alto IGM en la expresión de uno o los cinco genes, lo que indica una probabilidad de una mayor grado de respuesta a un AI por encima del tamoxifeno. Los ejemplos no limitantes de un AI incluyen los inhibidores no esteroideos tales como letrozol y anastrozol y los inhibidores esteroideos irreversibles tales como exemestano.

En más casos adicionales, la invención incluye composiciones y métodos para el uso de los niveles de expresión de uno o más de los cinco genes del IGM, o de los cinco, como un factor predictivo del resultado del tratamiento con quimioterapia. Opcionalmente, el ratio H:I también se somete a ensayo y se utiliza como una combinación con el gen o genes del IGM. El nivel o niveles de expresión de los genes pueden, de este modo, utilizarse para predecir la quimio-sensibilidad, tal como al paclitaxel/FAC (paclitaxel seguido de 5-fluorouracilo, doxorubicina y ciclofosfamida) o al taxol o a la terapia con antraciclina, como ejemplos no limitantes. Por tanto, la invención incluye la detección de la expresión de los cinco de estos genes, en la que la alta expresión del IGM es un indicador de una mayor probabilidad de una respuesta completa patológica (RCp) a la quimioterapia, tal como el tratamiento postoperatorio (intervención postquirúrgica) con paclitaxel/FAC como ejemplo no limitante. Como ejemplo no limitante, la detección puede ser de la expresión en una célula de cáncer de una muestra que contiene células preoperatorias utilizada para diagnosticar el cáncer en el sujeto. Como alternativa, el método puede comprender someter a ensayo cualquier subconjunto de estos cinco genes del IGM, hasta solo uno cualquiera de los genes, como factores predictivos de sensibilidad o resistencia a la quimioterapia.

El método puede incluir adicionalmente la identificación del sujeto como que tiene probabilidades, o no tiene probabilidades, de experimentar la RCp y, opcionalmente, puede incluir adicionalmente el ajuste de las modalidades de tratamiento para el sujeto para abordar el resultado esperado. Como ejemplo no limitante, la determinación de una baja probabilidad de RCp puede utilizarse para confirmar la idoneidad de, o para seleccionar, el tratamiento con quimioterapia, tal como paclitaxel/FAC. Como otro ejemplo no limitante, la determinación de una alta probabilidad de RCp puede utilizarse para confirmar la idoneidad de, o para seleccionar, la omisión de quimioterapia, tal como la omisión de paclitaxel/FAC, en favor de otras modalidades de tratamiento, tales como la mastectomía radical con inclusión de terapias postoperatorias, tales como la radiación.

La invención incluye adicionalmente composiciones y métodos para el uso de los niveles de expresión de uno o más de los cinco genes del IGM, o de los cinco, como un factor predictivo del grado de respuesta de un cáncer (sensibilidad) al tratamiento con radiación. Opcionalmente, el ratio H:I también se somete a ensayo y se utiliza como una combinación con el gen o genes del IGM. La alta expresión de IGM puede utilizarse, de este modo, para predecir si un paciente de cáncer de mama será sensible a un tratamiento con radiación, tal como una intervención postquirúrgica. Por tanto, la invención incluye la detección de la expresión de los cinco de estos genes, en la que la alta expresión del IGM es un indicador de la sensibilidad postoperatoria al tratamiento con radiación. Como ejemplo no limitante, la célula de cáncer puede ser una de una muestra histológica preoperatoria utilizada para diagnosticar el cáncer en el sujeto. Como alternativa, el método puede comprender el ensayo de cualquier subconjunto de los cinco genes del IGM, hasta solo uno cualquiera de los genes, como factores predictivos de sensibilidad o resistencia a la quimioterapia.

El método puede incluir adicionalmente la identificación del sujeto como que tiene probabilidades, o no tiene probabilidades, de ser sensible a la radioterapia después de la intervención quirúrgica, y, opcionalmente puede incluir adicionalmente el ajuste de las modalidades de tratamiento para el sujeto para abordar el resultado esperado. Como ejemplo no limitante, la determinación de una probabilidad de respuesta (sensibilidad) a la radiación postquirúrgica puede utilizarse para confirmar la idoneidad de, o para seleccionar, la radioterapia. Como otro ejemplo no limitante, la determinación de una baja probabilidad de respuesta (sensibilidad) a la radiación postquirúrgica puede utilizarse para confirmar la idoneidad de, o para seleccionar, la omisión de la radioterapia, opcionalmente en favor de la quimioterapia.

Breve descripción de las figuras

La Figura 1, parte A, ilustra la combinación del patrón de expresión de 5 genes en un solo índice (índice de grado molecular o IGM) a través de un análisis de componentes principales sin supervisión. El IGM se correlaciona fuertemente con el grado tumoral. La Figura 1, parte B, ilustra una agrupación basada en un modelo del IGM en todo el conjunto de datos, dando como resultado una distribución bimodal, con un punto de corte natural de aproximadamente 0. Este punto de corte clasificó correctamente la mayoría de los tumores de grado 1 y de grado 3 (precisión global del 89 %) y estratificó los tumores de grado 2 en dos grupos (el 59 % y el 41 % en el grupo del IGM bajo y alto, respectivamente). Los datos de probabilidad de supervivencia a lo largo de 12 años de los sujetos con tumores de grado 2 se representaron sobre la base de los valores de $IGM \leq 0$ y ≥ 0 en la parte C de la Figura 1 para ilustrar la capacidad de pronóstico del IGM.

La Figura 2 muestra una comparación del IGM y el IGG para la correlación con el grado tumoral y el resultado clínico. Las Partes A a C, ilustran una comparación del patrón de expresión de 5 genes con el IGG en la cohorte de Uppsala, mientras que las partes D a F corresponden a la cohorte de Estocolmo. Las Partes A y D son un análisis de la curva de característica operativa del receptor (ROC, por sus siglas en inglés) del IGM y el IGG para los discriminar los tumores de grado 1 y los de grado 3. Las partes B a F muestran las curvas de supervivencia de Kaplan-Meier que muestran la probabilidad de muerte específicamente por cáncer de mama de acuerdo con el estado del IGM o el IGG (alto frente a bajo).

La Figura 3 muestra las curvas de supervivencia de Kaplan-Meier de acuerdo con la firma de pronóstico de 76 genes o IGM en la cohorte TRANSBIG. Las Partes A y B son para todos los pacientes. Las partes C y D corresponden al grado tumoral RE+ subgrupo 1 o 2.

La Figura 4 muestra el IGM determinado mediante el ensayo de RT-PCR TaqMan™ en la cohorte MGH. La Parte A ilustra la correlación del IGM con el grado tumoral. Las partes B a D muestran los análisis de Kaplan-Meier de supervivencia sin metástasis a distancia de acuerdo con el IGM utilizando todos los pacientes (parte B), los con ganglios linfáticos negativos (parte C) o los pacientes con ganglios linfáticos positivos (Parte D).

La Figura 5 muestra los análisis de Kaplan-Meier de supervivencia sin metástasis a distancia de acuerdo con el IGM (parte A), el ratio H:I (parte B) o los tres grupos (riesgo bajo, intermedio y alto) generados por el peinado con el IGM y el ratio H:I (parte C) en la cohorte MGH, o los mismos tres grupos de riesgo en la cohorte de Oxford (parte D). La Figura 5E ilustra la interacción entre el IGM y el ratio H:I en la cohorte de la Tabla 1. Los pacientes con nodos negativos tratados con terapia endocrina o con terapia endocrina + quimioterapia (n = 93) de la cohorte se analizaron para determinar la interacción entre el IGM y el ratio H:I. El IGM era más consistente en la predicción de la metástasis a distancia en pacientes con alta HoxB13:IL17BR y, de forma similar, el ratio H:I era más consistente en pacientes con alto IGM.

La Figura 6 muestra la correlación del ratio H:I y el IGM con la expresión de RE, RP y HER2 como se determinó mediante RT-PCR en tiempo real en los pacientes con ganglios linfáticos negativos tratados con terapia endocrina de la cohorte de la Tabla 1. Eje X, grupos definidos por el ratio H:I, el IGM o su combinación. Eje Y, nivel de expresión relativo de RE, RP o HER2 como se indica.

La Figura 7 ilustra la interacción entre el IGM y el ratio H:I en la última cohorte. De forma similar a la cohorte de la Tabla 1, el IGM y el ratio H:I proporcionan información de pronóstico adicional el uno al otro. Los tumores con altos valores en ambos índices se asociaron a un pronóstico mucho peor que aquellos con solo un índice alto.

La Figura 8 ilustra la aplicación del IGM para identificar con mayor precisión las poblaciones de riesgo intermedio y bajo según los protocolos de St. Gallen en poblaciones de riesgo alto, intermedio y bajo.

La Figura 9 ilustra un resultado hipotético de un valor de IGM de 2,1 y su correlación con un riesgo del 19 % de reaparición del cáncer a los 5 años.

La Figura 10 muestra los análisis de curvas de Kaplan-Meier de la estratificación de los pacientes mediante el IGM de acuerdo con el tratamiento clínico o la falta del mismo. El panel A muestra los resultados con pacientes que no recibieron ningún tratamiento sistémico. El panel B muestra los resultados con pacientes que solo recibieron la terapia endocrina. RR = relación de riesgo a partir del análisis de regresión de Cox univariante y los valores de p son de un ensayo del orden logarítmico.

La Figura 11 muestra la capacidad predictiva del IGM para determinar la sensibilidad a la quimioterapia.

La Figura 12 muestra los análisis de curvas de Kaplan-Meier de la estratificación de los pacientes mediante el IGM de acuerdo con el estado pre o postmenopáusico. El panel A muestra los resultados con mujeres postmenopáusicas (edad ≥ 50). El panel B muestra los resultados con mujeres premenopáusicas (edad ≤ 50). RR = relación de riesgo a partir del análisis de regresión de Cox univariante y los valores de p son de un ensayo del orden logarítmico.

Descripción detallada de los modos de poner en práctica la invención

Definiciones de los términos como se utilizan en el presente documento:

Un "patrón" o "perfil" o "firma" de expresión génica se refiere a la expresión relativa de uno o más genes entre dos o más resultados clínicos, resultados de cáncer, reaparición del cáncer y/o resultados de supervivencia, que se correlaciona con la capacidad de distinguir entre dichos resultados. En algunos casos, el resultado es el de cáncer de mama.

Un "gen" es un polinucleótido que codifica un producto aislado, ya sea de naturaleza ARN o proteínica. Se aprecia que más de un polinucleótido puede ser capaz de codificar un producto aislado. El término incluye los alelos y los

polimorfismos de un gen que codifican el mismo producto o un análogo del mismo funcionalmente asociado (incluyendo la ganancia, la pérdida o la modulación de la función), sobre la base de la localización cromosómica y la capacidad de recombinarse durante la mitosis normal.

- 5 Los términos "correlacionar" o "correlación" o equivalentes de los mismos se refieren a una asociación entre la expresión de uno o más genes y un estado fisiológico de una célula a la exclusión de uno o más estados distintos como se identifica por el uso de los métodos como se describen en el presente documento. Un gen puede expresarse en un nivel superior o inferior y todavía correlacionarse con uno o más estados o resultados del cáncer.
- 10 Un "polinucleótido" es una forma polimérica de nucleótidos de cualquier longitud, ya sean ribonucleótidos o desoxirribonucleótidos. Este término se refiere solo a la estructura primaria de la molécula. Por tanto, este término incluye ADN y ARN bicatenarios y monocatenarios. También incluye tipos conocidos de modificaciones, incluyendo marcadores conocidos en la técnica, metilación, "caperuzas", sustitución de uno o más de los nucleótidos naturales por un análogo y modificaciones internucleotídicas tales como enlaces no cargados (por ejemplo, fosforotioatos, fosforoditioatos, etc.), así como formas no modificadas del polinucleótido.

El término "amplificar" se utiliza en sentido amplio para significar que la creación de un producto de amplificación puede hacerse enzimáticamente con ADN o ARN polimerasas. "Amplificación", como se utiliza en el presente documento, se refiere en general al proceso de producción de múltiples copias de una secuencia deseada, en particular las de una muestra. "Múltiples copias" significa al menos 2 copias. Una "copia" no significa necesariamente complementariedad de secuencia perfecta o identidad a la secuencia de la plantilla.

Por correspondiente se entiende que una molécula de ácido nucleico comparte una cantidad sustancial de identidad de secuencia con otra molécula de ácido nucleico. Cantidad sustancial significa al menos el 95 %, habitualmente al menos el 98 % y más habitualmente al menos el 99 %, y la identidad de secuencia se determina utilizando el algoritmo BLAST, como se describe en Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-410 (utilizando los ajustes predeterminados publicados, es decir, los parámetros $w = 4$, $t = 17$). Los métodos para la amplificación de ARNm se conocen en general en la técnica e incluyen la PCR de transcripción inversa (RT-PCR) y los descritos en la Solicitud de Patente de los EE.UU. 10/062.857 (presentada el 25 de octubre de 2001), así como en las Solicitudes de Patente Provisional de los EE.UU. 60/298.847 (presentada el 15 de junio de 2001) y 60/257.801 (presentada el 22 de diciembre de 2000). Otro método que puede utilizarse es la PCR cuantitativa (o Q-PCR). Como alternativa, el ARN puede marcarse directamente como el ADNc correspondiente mediante métodos conocidos en la técnica.

Una "micromatriz" es una matriz lineal o bidimensional de regiones preferentemente aisladas, cada una con un área definida, formada sobre la superficie de un soporte sólido tal como, pero no limitado a, vidrio, plástico o membrana sintética. La densidad de las regiones aisladas en una micromatriz se determina por el número total de polinucleótidos inmovilizados que se han de detectar en la superficie de un único soporte de fase sólida, preferentemente al menos aproximadamente $50/\text{cm}^2$, más preferentemente al menos aproximadamente $100/\text{cm}^2$, incluso más preferentemente al menos aproximadamente $500/\text{cm}^2$, pero preferentemente por debajo de aproximadamente $1.000/\text{cm}^2$. Preferentemente, las matrices contienen menos de aproximadamente 500, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 2500 o aproximadamente 3000 polinucleótidos inmovilizados en total. Como se utiliza en el presente documento, una micromatriz de ADN es una matriz de oligonucleótidos o polinucleótidos colocada sobre un chip u otras superficies utilizadas para hibridarse con los polinucleótidos amplificados o clonados de una muestra. Puesto que se conoce la posición de cada grupo particular de cebadores en la matriz, las identidades de una muestra de polinucleótidos pueden determinarse sobre la base de su unión a una posición particular en la micromatriz.

Debido a que la invención se basa en la identificación de genes que se sobre- o infraexpresan, una realización de la invención implica determinar la expresión mediante la hibridación de ARNm, o una versión amplificada o clonada del mismo, de una célula de muestra con un polinucleótido que es único para una secuencia génica particular. Los polinucleótidos preferidos de este tipo contienen al menos aproximadamente 20, al menos aproximadamente 22, al menos aproximadamente 24, al menos aproximadamente 26, al menos aproximadamente 28, al menos aproximadamente 30 o al menos aproximadamente 32 pares de bases consecutivas de una secuencia génica que no se encuentran en otras secuencias génicas. El término "aproximadamente" como se ha utilizado en la frase anterior se refiere a un aumento o disminución del 1 desde el valor numérico indicado. Incluso se prefieren más los polinucleótidos de al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350 o al menos o aproximadamente 400 pares de bases de una secuencia génica que no se encuentran en otras secuencias génicas. El término "aproximadamente" como se ha utilizado en la frase anterior se refiere a un aumento o disminución del 10 % del valor numérico indicado. Dichos polinucleótidos también pueden denominarse sondas polinucleotídicas que son capaces de hibridarse con secuencias de los genes, o partes únicas de los mismos, descritas en el presente documento. Preferentemente, las secuencias son las del ARNm codificado por los genes, el ADNc correspondiente a dichos ARNm, y/o versiones amplificadas de dichas secuencias. En realizaciones preferidas de la divulgación, las sondas polinucleotídicas se inmovilizan sobre una matriz, otros dispositivos, o en puntos individuales que localizan las sondas.

- En otra realización de la invención, la totalidad o parte de una secuencia desvelada puede amplificarse y detectarse por métodos tales como la reacción en cadena de la polimerasa (PCR) y variaciones de la misma, tales como, pero no limitadas a, la PCR cuantitativa (Q-PCR), la PCR de transcripción inversa (RT-PCR) y la PCR en tiempo real, opcionalmente RT-PCR en tiempo real. Dichos métodos serían utilizar uno o dos cebadores que son
- 5 complementarios a porciones de una secuencia desvelada, donde los cebadores se utilizan para la síntesis de ácido nucleico cebado. Los ácidos nucleicos recién sintetizados están opcionalmente marcados y pueden detectarse directamente o mediante hibridación con un polinucleótido de la invención. Los ácidos nucleicos recién sintetizados pueden ponerse en contacto con polinucleótidos (que contienen secuencias) de la invención en condiciones que permitan su hibridación.
- 10 Como alternativa, y en otra realización de la invención, la expresión génica puede determinarse mediante el análisis de la proteína expresada en una muestra celular de interés mediante el uso de uno o más anticuerpos específicos para uno o más epítomos de los productos génicos individuales (proteínas) en dicha muestra celular. Dichos anticuerpos se marcan preferentemente para permitir su detección fácil después de la unión al producto génico.
- 15 El término "marcador" se refiere a una composición capaz de producir una señal detectable indicativa de la presencia de la molécula marcada. Los marcadores adecuados incluyen radioisótopos, cromóforos nucleotídicos, enzimas, sustratos, moléculas fluorescentes, restos quimioluminiscentes, partículas magnéticas, restos bioluminiscentes y similares. Como tal, un marcador es cualquier composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunquímicos, eléctricos, ópticos o químicos.
- 20 El término "soporte" se refiere a soportes convencionales tales como perlas, partículas, tiras reactivas, fibras, filtros, membranas y soportes de silano o silicato tales como portaobjetos de vidrio.
- Como se utiliza en el presente documento, una "muestra de tejido canceroso" o "muestra de células cancerosas" se refieren a una célula que contiene la muestra de tejido aislado de un individuo que padece el cáncer correspondiente. La muestra puede ser de material retirado a través de un procedimiento quirúrgico, tal como una
- 25 biopsia. Dichas muestras son aislados primarios (a diferencia de las células cultivadas) y pueden recogerse por cualquier medio adecuado reconocido en la técnica. En algunas realizaciones, la "muestra" puede recogerse por un método no invasivo, incluyendo, pero no limitado a, la abrasión o la aspiración con aguja fina.
- 30 Una "muestra de tejido mamario" o "muestra de células mamarias" se refiere a una muestra de tejido mamario o fluido aislado de un individuo que se sospecha que padece, o está en riesgo de desarrollar, cáncer de mama. Dichas muestras son aislados primarios (a diferencia de las células cultivadas) y pueden recogerse por cualquier medio no invasivo, incluyendo, pero no limitado a, el lavado ductal, la aspiración con aguja fina, la biopsia con aguja, los
- 35 dispositivos y métodos descritos en la Patente de los EE.UU. N.º 6.328.709 o cualquier otro medio adecuado reconocido en la técnica. Como alternativa, la "muestra" puede recogerse por un método invasivo, incluyendo, pero no limitado a, la biopsia quirúrgica.
- "Expresión" y "expresión génica" incluyen la transcripción y/o traducción de material de ácido nucleico. Por supuesto, el término también puede limitarse, si así se indica, para referirse únicamente a la transcripción de ácidos nucleicos.
- 40 Como se utiliza en el presente documento, la expresión "que comprende" y sus afines se utilizan en su sentido inclusivo; es decir, equivalente a la expresión "que incluye" y sus correspondientes afines.
- 45 Las condiciones que "permiten" que se produzca un acontecimiento o las condiciones que son "adecuadas" para que un acontecimiento se produzca, tal como la hibridación, la extensión de la cadena y similares, o las condiciones "adecuadas" son condiciones que no impiden que dichos acontecimientos se produzcan. Por tanto, estas condiciones permiten, potencian, facilitan y/o propician el acontecimiento. Dichas condiciones, conocidas en la
- 50 técnica y descritas en el presente documento, dependen de, por ejemplo, la naturaleza de la secuencia nucleotídica, la temperatura y la condiciones tamponantes. Estas condiciones también dependen de qué acontecimiento se desea, tal como la hibridación, la escisión, la extensión o la transcripción de la cadena.
- La "mutación" de la secuencia, como se utiliza en el presente documento, se refiere a cualquier alteración de la secuencia en la secuencia de un gen desvelado en el presente documento de interés en comparación con una
- 55 secuencia de referencia. Una mutación de secuencia incluye cambios de un único nucleótido o alteraciones de más de un nucleótido en una secuencia, debido a mecanismos tales como la sustitución, la delección o la inserción. El polimorfismo de nucleótido único (PNU) también es una mutación de secuencia como se utiliza en el presente documento. Debido a que la presente invención se basa en el nivel relativo de expresión génica, las mutaciones en regiones no codificantes de genes como se desvelan en el presente documento también pueden ensayarse en la
- 60 práctica de la invención.
- "Detección" incluye cualquier medio de detección, incluyendo la detección directa e indirecta de la expresión génica y los cambios en la misma. Por ejemplo, pueden observarse "detectablemente menos" productos directamente o indirectamente y el término indica cualquier reducción (incluyendo la ausencia de señal detectable). De forma
- 65 similar, "detectablemente más" producto significa cualquier aumento, observado ya sea directamente o indirectamente.

Los aumentos y disminuciones en la expresión de las secuencias desveladas se definen en los siguientes términos basados en los cambios porcentuales o en número de veces por encima de la expresión en las células normales. Los aumentos pueden ser del 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180 o 200 % en relación con los niveles de expresión en las células normales. Como alternativa, los aumentos en número de veces pueden ser de 1, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 o 10 veces por encima de los niveles de expresión en las células normales. Las disminuciones pueden ser del 10, 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 99 o 100 % en relación con los niveles de expresión en las células normales.

A menos que se defina otra cosa, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que se entiende frecuentemente por un experto habitual en la técnica a la que pertenece la presente invención.

General

En la presente invención, a través de enfoques basados tanto en los datos como en el conocimiento, una firma de grado tumoral de 5 genes (IGM) se desarrolló y se implementó en un ensayo de RT-PCR consistente. Una característica importante del IGM es que su cálculo no implica una ponderación compleja basada en el resultado clínico. En cambio, es una correlación molecular del grado tumoral y su capacidad de pronóstico deriva de este último (denominado enfoque "ascendente"). La ventaja del IGM sobre el grado tumoral histológico es doble. En primer lugar, como el IGG, clasifica los tumores de grado 2 ya sea como similares al grado 1 o similares al grado 3, eliminando la mayor parte de la ambigüedad de la clasificación del tumor patológico. En segundo lugar, y debido a que un ensayo basado en RT-PCR puede normalizarse en el laboratorio clínico, también elimina la subjetividad y la variabilidad inter/intraobservador asociada a la clasificación patológica.

Los resultados divulgados también muestran que la precisión de pronóstico del IGM puede aumentarse al considerar también el ratio H:I y viceversa, lo que señala un algoritmo simple que estratifica a los pacientes en tres grupos de riesgo. El IGM y el ratio H:I parecen representar dos módulos de pronóstico distintos en el cáncer de mama, como se señaló por la observación de que el ratio H:I, pero no el IGM, se asocia a la señalización estrogénica.

Más allá de sus capacidades de pronóstico, el IGM y el ratio H:I también son factores predictivos potenciales para el beneficio terapéutico de la quimioterapia y la terapia endocrina, respectivamente. Un alto grado tumoral o índice mitótico predicen el beneficio de la quimioterapia en pacientes con cáncer de mama con ganglios negativos. De forma similar, se ha demostrado que el grupo de proliferación de los genes en el algoritmo del índice de recurrencia predice el beneficio de la quimioterapia en pacientes RE+ ganglios negativos. De hecho, un IGM alto predice la respuesta patológica completa en pacientes con cáncer de mama RE+ tratados con paclitaxel preoperatorio seguido de 5-fluorouracilo, doxorubicina y ciclofosfamida.

Dos estudios recientes del ratio H:I han demostrado su potencial como un nuevo biomarcador del grado de respuesta endocrina más allá de los receptores de estrógeno y progesterona. En un estudio de cáncer de mama recurrente, una relación H:I baja estaba fuertemente asociada a la respuesta a la terapia de primera línea con tamoxifeno. De forma similar, en un análisis de muestras tumorales de un ensayo clínico aleatorizado prospectivo comparando 2 años frente a 5 años de tratamiento con tamoxifeno, un bajo HoxB13 o una relación H:I baja predijeron significativamente el beneficio de la terapia prolongada con tamoxifeno. Estos resultados son coherentes con la observación de que el estrógeno regula negativamente el HoxB13 y regula positivamente la expresión de IL17BR. Por tanto, en los tumores RE+, un HoxB13 o una relación H:I altos pueden considerarse como un marcador de señalización estrogénica disfuncional.

Los papeles dobles del IGM y el ratio H:I son especialmente relevantes en el contexto de las últimas directrices del consenso de St. Gallen (2005) para la selección del tratamiento para el cáncer de mama en estadio temprano. Las directrices de St. Gallen clasifican a los pacientes con cáncer de mama RE+ con ganglios negativos en grupos de riesgo bajo e intermedio, perteneciendo la mayoría al segundo. Una decisión importante es si mantener el tratamiento con quimioterapia para algunos de los pacientes del grupo de riesgo intermedio, una cuestión abordada por dos nuevos ensayos clínicos prospectivos. En la cohorte de la Tabla 1 descrita en el presente documento, la aplicación de las directrices de St. Gallen dio como resultado la clasificación del 86 % de los pacientes en el grupo de riesgo intermedio, que podrían volver a clasificarse como de bajo (43 %), intermedio (26 %) o alto (31 %) riesgo utilizando el IGM y el ratio H:I.

La excelente probabilidad de supervivencia sin enfermedad de los pacientes de bajo riesgo señala que potencialmente podría ahorrarseles la quimioterapia tóxica sin comprometer su pronóstico, mientras que deberían añadirse pautas de quimioterapia más intensas o nuevos agentes terapéuticos al grupo de alto riesgo. Por tanto, la estratificación del riesgo basada en el IGM y el ratio H:I y sus respectivas capacidades predictivas podrían añadirse a las directrices existentes para equilibrar mejor la relación riesgo-beneficio de las modalidades de tratamiento actuales.

Por tanto, la presente invención incluye un IGM validado como un potente factor de pronóstico en el cáncer de

mama RE+. Además, el IGM y el ratio H:I pueden combinarse para proporcionar información de pronóstico más precisa que cualquiera de ellos solo. La identificación de un subgrupo de pacientes con muy mal resultado utilizando estos dos biomarcadores facilita el diseño de ensayos clínicos para dirigirse a los cánceres tanto con IGM alto como relación H:I alta.

5

IGM

Los genes divulgados como se expresa en correlación con los grados tumorales particulares proporcionan la capacidad de enfocar el análisis de la expresión génica a solo aquellos genes que contribuyen a la capacidad de identificar un sujeto como que tiene probabilidades de tener un pronóstico particular, o un resultado clínico, con respecto a otro. La expresión de otros genes en una célula cancerosa sería relativamente incapaz de proporcionar información de interés y, por tanto, de ayudar en estas discriminaciones.

10

Para determinar los niveles de expresión de genes en la práctica de la presente invención, puede utilizarse cualquier método conocido en la técnica. En algunas realizaciones, se utiliza la expresión basada en la detección de ARN que se hibrida con los genes identificados y divulgados en el presente documento. Esto se realiza fácilmente mediante cualquier método de detección o amplificación + detección de ARN, conocido o reconocido como equivalente en la técnica tal como, pero no limitado a, transcripción inversa-PCR, los métodos divulgados en la solicitud de patente de los EE.UU. N.º de serie 10/062.857 (presentada el 25 de octubre de 2001), así como la solicitud de patente provisional de los EE.UU. 60/298.847 (presentada el 15 de junio de 2001) y 60/257.801 (presentada el 22 de diciembre de 2000) y métodos para detectar la presencia o ausencia de secuencias estabilizadoras o desestabilizadoras del ARN.

15

20

Como alternativa, puede utilizarse la expresión basada en la detección del estado del ADN. La detección del ADN de un gen identificado como metilado o eliminado puede utilizarse para genes que tienen una menor expresión. Esto puede realizarse fácilmente por métodos basados en la PCR conocidos en la técnica, incluyendo, pero no limitados a, Q-PCR. Por el contrario, la detección del ADN de un gen identificado como amplificado puede utilizarse para los genes que tienen una mayor expresión en correlación con un resultado de cáncer de mama particular. Esto puede realizarse fácilmente por métodos basados en la PCR, hibridación *in situ* fluorescente (FISH) e hibridación cromosómica *in situ* (CISH) conocidos en la técnica.

25

30

También puede utilizarse la expresión basada en la detección de una presencia, aumento o disminución de los niveles o actividad de proteínas. La detección puede realizarse por cualquier método basado en la inmunohistoquímica (IHC), basado en sangre (especialmente para proteínas secretadas), basado en anticuerpos (incluyendo autoanticuerpos contra la proteína), basado en células de exfoliado (del cáncer), basado en espectroscopia de masas y basado en imágenes (incluyendo el uso de ligando marcado), conocido en la técnica y reconocido como apropiado para la detección de la proteína. Los métodos basados en anticuerpos e imágenes son, adicionalmente, útiles para la localización de tumores después de la determinación del cáncer mediante el uso de células obtenidas por un procedimiento no invasivo (tal como el lavado ductal o la aspiración con aguja fina), donde no se conoce el origen de las células cancerosas. Pueden utilizarse un anticuerpo o ligando marcados para localizar el carcinoma o carcinomas en un paciente.

35

40

Una realización utilizando un ensayo basado en ácidos nucleicos para determinar la expresión es mediante la inmovilización de una o más secuencias de los genes identificados en el presente documento sobre un soporte sólido, incluyendo, pero no limitado a, un sustrato sólido como una matriz o a perlas o tecnología basada en perlas como se conocen en la técnica. Como alternativa, también pueden utilizarse ensayos de expresión basados en solución conocidos en la técnica.

45

El gen o genes inmovilizados pueden estar en forma de polinucleótidos que son únicos o de otra manera específicos para el gen o genes de manera que el polinucleótido sería capaz de hibridarse con un ADN o ARN correspondiente al gen o genes. Estos polinucleótidos pueden ser la longitud completa del gen o genes o pueden ser secuencias cortas de los genes (hasta un nucleótido más cortas que la secuencia de longitud completa conocida en la técnica por delección del extremo 5' o 3' de la secuencia) que están opcionalmente mínimamente interrumpidas (tal como por desapareamientos o pares de bases no complementarias insertadas) de manera que la hibridación con un ADN o ARN correspondiente al gen o genes no se ve afectada. En algunos casos, los polinucleótidos utilizados son del extremo 3' del gen, tales como de aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100 o aproximadamente 50 nucleótidos desde la señal de poliadenilación o el sitio de poliadenilación de un gen o secuencia expresada. Los polinucleótidos que contienen mutaciones relacionadas con las secuencias de los genes divulgados también pueden utilizarse siempre que la presencia de las mutaciones todavía permita la hibridación para producir una señal detectable.

50

55

60

El gen o genes inmovilizados pueden utilizarse para determinar el estado de muestras de ácidos nucleicos preparadas a partir de células de cáncer de muestra, o de mama, para las que el resultado del sujeto de la muestra (por ejemplo, el paciente del que se obtiene la muestra) no se conoce o para la confirmación de un resultado que ya se ha asignado al sujeto de la muestra. Sin limitar la invención, una célula de este tipo puede ser de un paciente con cáncer de mama RE+. El polinucleótido o polinucleótidos inmovilizados solo necesitan ser suficientes para hibridarse

65

específicamente con las moléculas de ácido nucleico correspondientes derivadas de la muestra en condiciones adecuadas.

La invención se basa en parte en el descubrimiento de un factor de pronóstico y predictivo basado en la expresión génica del resultado clínico y el grado tumoral, tal como el que utiliza muestras de cáncer de tejidos FFIP, muestras congeladas o muestras frescas. Los niveles de expresión de estos genes se correlacionan con el grado tumoral y los resultados clínicos como se describe en el presente documento, así como la determinación del pronóstico de un sujeto. Los genes identificados tienen papeles en el ciclo celular y notificaron una expresión máxima de la siguiente manera:

Gen	Pico de Expresión	Papel en el ciclo celular
BUB1B	G2/M	Punto de control del ensamblaje del huso mitótico
CENPA	G2/M	Ensamblaje del centrómero
NEK2	G2/M	Duplicación del centrosoma
RACGAP1	No determinado	Iniciación de la citocinesis
RRM2	S	Replicación del ADN

Las secuencias de estos genes se han notificado y caracterizado anteriormente en el campo. Por ejemplo, el gen BUB1B humano se identifica mediante Unigene Hs.631699, que se caracteriza por 273 secuencias correspondientes, el 6 de septiembre de 2007.

En la misma fecha, los genes CENPA, NEK2, RACGAP1 y RRM2 humanos se identifican mediante Hs. 1594 (con 129 secuencias correspondientes), Hs. 153704 (con 221 secuencias correspondientes), Hs. 696319 (con 349 secuencias correspondientes) y Hs. 226390 (con 1348 secuencias correspondientes), respectivamente. Las secuencias correspondientes a cada uno de estos identificadores Unigene se incorporan en el presente documento por referencia como si se expusiera con todo detalle y pueden utilizarse en la práctica de la invención por el experto en la materia según se considere apropiado.

Se divulgan dos de doce secuencias representativas de ARNm de BUB1B identificadas mediante Unigene Hs.631699 en el listado de secuencias; se desvelan dos de once secuencias representativas de ARNm de CENPA identificadas mediante Hs. 1594 en el listado de secuencias; se desvelan dos de diecisiete secuencias representativas de ARNm de NEK2 identificadas mediante Hs. 15704 en el listado de secuencias; se desvelan dos de quince secuencias representativas de ARNm de RACGAP1 identificadas mediante Hs. 696319 en el listado de secuencias; y se desvelan dos de 24 secuencias representativas de ARNm de RRM2 identificadas mediante Hs. 226390 en el listado de secuencias. Las secuencias desveladas en el listado son no limitantes para la práctica de la invención desvelada, pero se proporcionan como pruebas del conocimiento sustancial en el campo con respecto a las secuencias que son los genes divulgados. Adicionalmente, el experto es plenamente capaz de alinear cualesquiera dos o más de las secuencias conocidas para cada uno de estos genes para identificar áreas de identidad o cambios conservados como regiones que identifican cada uno de estos genes.

Las secuencias de los mismos genes también se han identificado y caracterizado a partir de otras especies animales. De este modo, el experto en el campo es claramente consciente de cómo identificar los genes divulgados con respecto a otros genes animales. El experto en la materia también puede comparar opcionalmente las secuencias conocidas de los genes divulgados de diferentes fuentes animales para identificar regiones y secuencias conservadas únicas para estos genes con respecto a otros genes.

De forma similar, el uso de STK15, Survivina, Ciclina B1 y MYBL2 como se describe en el presente documento se apoya en los informes anteriores con respecto a estos genes y secuencias representativas de cada uno de estos genes conocidos para el experto en la materia.

Como se apreciará por los expertos en la materia, algunas de las secuencias correspondientes indicadas anteriormente incluyen tramos 3' poli A (o poli T en la cadena complementaria) que no contribuyen a la singularidad de las secuencias desveladas. La invención puede, por tanto, ponerse en práctica con secuencias que carecen de tramos 3' poli A (o poli T). La singularidad de las secuencias desveladas se refiere a las porciones o la totalidad de las secuencias que solo se encuentran en los ácidos nucleicos de los genes divulgados, incluyendo secuencias únicas que se encuentran en la porción 3' no traducida de los genes. Las secuencias únicas preferidas para la práctica de la invención son las que contribuyen a las secuencias de consenso para cada uno de los tres conjuntos de manera que las secuencias únicas serán útiles en la detección de la expresión en una diversidad de individuos en lugar de ser específicas para un polimorfismo presente en algunos individuos. Como alternativa, pueden utilizarse secuencias únicas para un individuo o una subpoblación. Las secuencias únicas preferidas son preferentemente de las longitudes de los polinucleótidos de la invención como se analiza en el presente documento.

Para determinar los niveles de expresión (aumentados o disminuidos) de las secuencias descritas anteriormente en la práctica de la invención, puede utilizarse cualquier método conocido en la técnica. En una realización de la invención, se utiliza la expresión basada en la detección de ARN que se hibrida con polinucleótidos que contienen

las secuencias descritas anteriormente. Esto se realiza fácilmente mediante cualquier método conocido de detección o de amplificación + detección de ARN conocido o reconocido como equivalente en la técnica tal como, pero no limitado a, la transcripción inversa-PCR (opcionalmente la PCR en tiempo real), los métodos divulgados en la solicitud de patente de los EE.UU. N.º de serie 10/062.857, titulada "*Nucleic Acid Amplificación*", presentada el 25 de octubre de 2001, así como la Solicitud de Patente Provisional de los EE.UU. 60/298.847 (presentada el 15 de junio de 2001) y 60/257.801 (presentada el 22 de diciembre de 2000), los métodos divulgados en la patente de los EE.UU. N.º 6.291.170, y la PCR cuantitativa. También pueden utilizarse métodos de identificación de un aumento de la estabilidad del ARN (que da como resultado una observación de aumento de la expresión) o de una disminución de la estabilidad del ARN (que da como resultado una observación de disminución de la expresión). Estos métodos incluyen la detección de secuencias que aumentan o disminuyen la estabilidad de los ARNm que contienen las secuencias de los genes. Estos métodos también incluyen la detección de un aumento de la degradación del ARNm.

En algunas realizaciones de la invención, se utilizan polinucleótidos que tienen secuencias presentes en las regiones no traducidas y/o no codificantes 3' de las secuencias desveladas anteriormente para detectar los niveles de expresión de las secuencias génicas en las células de cáncer o mamarias. Dichos polinucleótidos pueden contener opcionalmente secuencias encontradas en las porciones 3' de las regiones codificantes de las secuencias desveladas anteriormente. Los polinucleótidos que contienen una combinación de secuencias de las regiones codificantes y 3' no codificantes tienen preferentemente las secuencias dispuestas de forma contigua, sin la intervención de secuencia o secuencias heterólogas.

Como alternativa, la invención puede ser puesta en práctica con polinucleótidos que tienen secuencias presentes en regiones 5' no traducidas y/o no codificantes de las secuencias génicas en las células de cáncer o mamarias, para detectar sus niveles de expresión. Dichos polinucleótidos pueden contener opcionalmente secuencias encontradas en las porciones 5' de las regiones codificantes. Los polinucleótidos que contienen una combinación de secuencias de las regiones codificantes y 5' no codificantes tienen preferentemente las secuencias dispuestas de forma contigua, sin la intervención de secuencia o secuencias heterólogas. La invención también puede ser puesta en práctica con secuencias presentes en las regiones codificantes de las secuencias génicas desveladas.

Los polinucleótidos no limitantes contienen secuencias de regiones 3' o 5' no traducidas y/o no codificantes de al menos aproximadamente 20, al menos aproximadamente 22, al menos aproximadamente 24, al menos aproximadamente 26, al menos aproximadamente 28, al menos aproximadamente 30, al menos aproximadamente 32, al menos aproximadamente 34, al menos aproximadamente 36, al menos aproximadamente 38, al menos aproximadamente 40, al menos aproximadamente 42, al menos aproximadamente 44 o al menos aproximadamente 46 nucleótidos consecutivos. El término "aproximadamente" como se utiliza en la frase anterior se refiere a un aumento o disminución de 1 del valor numérico indicado. Incluso se prefieren más los polinucleótidos que contienen secuencias de al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350 o al menos o aproximadamente 400 nucleótidos consecutivos. El término "aproximadamente" como se utiliza en la frase anterior se refiere a un aumento o disminución del 10 % del valor numérico indicado.

Las secuencias del extremo 3' o 5' de las regiones codificantes descritas anteriormente como se encuentran en los polinucleótidos de la invención son de las mismas longitudes que las descritas anteriormente, excepto por que, naturalmente, estarían limitadas por la longitud de la región codificante. El extremo 3' de una región codificante puede incluir secuencias de hasta la mitad 3' de la región codificante. A la inversa, el extremo 5' de una región codificante puede incluir secuencias de hasta la mitad 5' de la región codificante. Por supuesto, las secuencias descritas anteriormente o las regiones y polinucleótidos codificantes que contienen porciones de las mismas, pueden utilizarse en su totalidad.

Los polinucleótidos que combinan las secuencias de una región 3' no traducida y/o no codificante y el extremo 3' asociado de la región codificante puede ser al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350, al menos o aproximadamente 400 nucleótidos consecutivos. Preferentemente, los polinucleótidos utilizados son del extremo 3' del gen, tales como en aproximadamente 350, aproximadamente 300, aproximadamente 250, aproximadamente 200, aproximadamente 150, aproximadamente 100 o aproximadamente 50 nucleótidos de la señal de poliadenilación o sitio de poliadenilación de un gen o una secuencia expresada. También pueden utilizarse polinucleótidos que contienen mutaciones relativas a las secuencias de los genes divulgados siempre que la presencia de las mutaciones todavía permita la hibridación para producir una señal detectable.

En otra realización de la invención, pueden utilizarse polinucleótidos que contienen deleciones de nucleótidos del extremo 5' y/o 3' de las secuencias desveladas anteriormente. Las deleciones son preferentemente de 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-175 o 175-200 nucleótidos del extremo 5' y/o 3', aunque el alcance de las deleciones naturalmente estaría limitado por la longitud de las secuencias desveladas y la necesidad de tener la capacidad de utilizar los polinucleótidos para la detección de niveles de expresión.

Otros polinucleótidos de la invención del extremo 3' de las secuencias desveladas anteriormente incluyen los de los cebadores y sondas opcionales para la PCR cuantitativa. En algunas realizaciones, los cebadores y sondas son los que amplifican una región de menos de aproximadamente 350, menos de aproximadamente 300, menos de
 5 aproximadamente 250, menos de aproximadamente 200, menos de aproximadamente 150, menos de aproximadamente 100 o menos de aproximadamente 50 nucleótidos de la señal de poliadenilación o sitio de poliadenilación de un gen o secuencia expresada.

En otras realizaciones más de la invención, pueden utilizarse polinucleótidos que contienen porciones de las
 10 secuencias desveladas anteriormente incluyendo el extremo 3'. Dichos polinucleótidos podrían contener al menos o aproximadamente 50, al menos o aproximadamente 100, al menos o aproximadamente 150, al menos o aproximadamente 200, al menos o aproximadamente 250, al menos o aproximadamente 300, al menos o aproximadamente 350 o al menos o aproximadamente 400 nucleótidos consecutivos desde el extremo 3' de las secuencias desveladas.

La invención también incluye polinucleótidos utilizados para detectar la expresión génica en células mamarias. Los
 15 polinucleótidos pueden comprender un polinucleótido más corto que consiste en secuencias encontradas en los genes anteriores en combinación con secuencias heterólogas que no se encuentran de forma natural en combinación con las secuencias. Los ejemplos no limitantes incluyen secuencias cortas de vectores de clonación o
 20 presentes en los fragmentos de restricción utilizados para preparar sondas o cebadores marcados, como se describe en el presente documento.

Métodos

Como se describe en el presente documento, la invención incluye la identidad de los genes, cuya expresión puede
 25 utilizarse para proporcionar información de pronóstico relacionada con el cáncer. En particular, los niveles de expresión de estos genes pueden utilizarse en relación con el cáncer de mama. En algunos métodos, el perfil de expresión génica se correlaciona con (y de este modo son capaces de discriminar entre) pacientes con buenos o malos resultados de reparación del cáncer y/o de supervivencia. En otras realizaciones, la invención incluye un
 30 método para comparar la expresión génica en una muestra de células cancerosas de un paciente con el perfil de expresión génica para determinar el resultado probable clínico o de tratamiento para el paciente, o el resultado biológico natural, en ausencia de intervención. Estas realizaciones de la invención pueden utilizarse ventajosamente para satisfacer una necesidad de diagnóstico no satisfecha, importante, de la capacidad de predecir si es probable que un paciente se beneficie de un tipo de tratamiento dado o si un paciente estará mejor con otro tipo de
 35 tratamiento. Por ejemplo, un valor bajo de el ratio H:I está fuertemente asociado a la respuesta a la terapia de primera línea con tamoxifeno. Y un análisis de muestras tumorales de un ensayo clínico aleatorizado prospectivo comparando 2 años frente a 5 años de tratamiento con tamoxifeno indica que una baja HOXB 13 o una baja HOXB 13:IL17BR predice significativamente el beneficio de la terapia prolongada con tamoxifeno.

De forma similar, la capacidad del IGM para predecir la presencia de tumores de grado I frente a tumores de grado
 40 III permite al médico del campo pertinente seleccionar la terapia apropiada para los cánceres de estos dos grados. El IGM no solo confirma las clasificaciones de grado I y de grado III realizadas por otros medios, sino que puede, en combinación con el ratio H:I, clasificar con mayor precisión los tumores de riesgo bajo, intermedio o alto que se han clasificado incorrectamente por otros métodos. Esto se ilustra en la Figura 8, donde la estratificación basada en el
 45 protocolo de 2005 de St. Gallen se corrige de manera significativa mediante el uso del IGM y el ratio H:I divulgados.

De este modo, la invención incluye un método de identificación de un paciente, de una población de pacientes con
 50 células cancerosas, como pertenecientes a una subpoblación de pacientes con un mejor pronóstico o una subpoblación con un mal pronóstico. La subpoblación con un mejor pronóstico es similar a los sujetos identificados con un tumor de grado I en comparación con una subpoblación de pacientes con peor pronóstico, similar a los sujetos identificados con un tumor de grado III. Por supuesto, los métodos divulgados no son necesariamente perfectos en su aplicación y es posible que un paciente dado se identifique como que tiene un grado tumoral "intermedio" entre el de los grados I y III. En cuyo caso, el facultativo experto podría tratar el sujeto en consecuencia.

No obstante, la invención proporciona un medio no subjetivo para la identificación de pacientes con tumores de
 55 grado I, intermedio o de grado III, identificación que puede utilizarse para el beneficio de un paciente por el facultativo experto. Es importante destacar que los métodos divulgados pueden clasificar los tumores de grado "intermedio" por otros métodos en estado de grado I o de grado III. Esto proporciona un enorme beneficio para la subpoblación de pacientes correspondiente, que de otro modo se habría tratado como si tuviera tumores de grado "intermedio". De este modo, en algunas realizaciones, se proporciona un método de reducción del número de
 60 clasificaciones de grado "intermedio" mediante el uso del IGM de 5 genes desvelado.

De este modo, la invención incluye un método de determinación del pronóstico y/o los resultados de supervivencia
 65 mediante el ensayo de los patrones de expresión divulgados en el presente documento. De este modo, donde ha podido utilizarse anteriormente la interpretación subjetiva para determinar el pronóstico y/o el tratamiento de pacientes con cáncer, la presente invención proporciona patrones de expresión génica objetivos, que pueden

utilizarse solos o en combinación con criterios subjetivos para proporcionar una evaluación más precisa de los resultados del paciente, incluyendo la supervivencia y la reaparición del cáncer. En algunos casos, el ensayo incluye detectar el nivel de expresión de Bub1B, en el que el nivel de expresión se correlaciona con un tumor de grado I o de grado III.

5 Los genes divulgados se identifican como correlacionados con el grado tumoral y los resultados clínicos de manera que los niveles de su expresión son pertinentes para una determinación de los protocolos de tratamiento de un paciente. De este modo, en algunas realizaciones, la invención proporciona un método para determinar el
10 tratamiento terapéutico para un paciente con cáncer mediante la determinación del pronóstico para dicho paciente mediante el ensayo de una muestra de células de cáncer de dicho paciente para los niveles de expresión descritos en el presente documento para determinar el grado tumoral y la selección de un tratamiento para un paciente con un tumor de dicho grado. En algunos casos, el ensayo incluye detectar el nivel de expresión de Bub1B, en el que el nivel de expresión se correlaciona con un tumor de grado I o de grado III.

15 En un conjunto de realizaciones, un método de la invención puede incluir el ensayo de una muestra de células cancerosas de un sujeto que padece cáncer para el nivel de expresión de Bub1B en el que el nivel de expresión clasifica el cáncer como correspondiente a un tumor de grado I o de grado III, o identifica al sujeto como que tiene un pronóstico de reaparición del cáncer probable, o predice el grado de respuesta del sujeto al tratamiento con terapia endocrina, quimioterapia o radioterapia. El ensayo puede incluir medir o detectar o determinar el nivel de expresión
20 del gen en cualquier medio adecuado descrito en el presente documento o conocido por el experto en la materia. En muchos casos, el cáncer es el cáncer de mama y el sujeto es un paciente humano. Adicionalmente, las células cancerosas pueden ser las de un tumor y/o de un sujeto con ganglio negativo (ganglios linfáticos negativos para el cáncer) o ganglio positivo (ganglios linfáticos positivos para el cáncer).

25 Por supuesto, el método puede ponerse en práctica junto con el ensayo de la expresión de uno o más de los otros cuatro genes del IGM, en el que los niveles de expresión de los genes utilizados en combinación se utilizan para clasificar, identificar o predecir, como se proporcionan por el método. El nivel requerido de nivel de expresión puede ser el que se identifica mediante los métodos descritos en el presente documento para los genes utilizados. Adicionalmente, el ensayo puede incluir la preparación de ARN a partir de la muestra, opcionalmente para el uso en
30 la PCR (reacción en cadena de la polimerasa) u otra metodología de análisis como se describe en el presente documento. La metodología de la PCR es opcionalmente RT-PCR (transcripción inversa-PCR) o la PCR cuantitativa, como la RT-PCR en tiempo real. Como alternativa, el ensayo puede realizarse mediante el uso de una matriz, tal como una micromatriz como se conoce en el campo pertinente. Opcionalmente, la muestra de las células cancerosas se disecciona a partir de tejido retirado u obtenido de dicho sujeto. Como se describe en el presente documento, puede utilizarse una diversidad de tipos de muestras, incluyendo una muestra fijada con formol incluida
35 en parafina (FFIP) como ejemplo no limitante. Y como se describe en el presente documento, el método puede incluir el ensayo o la determinación del ratio H:I (relación de los niveles de expresión de HoxB13 y IL17BR) en la muestra, como se desvela en el presente documento.

40 A modo de ejemplo no limitante, los cinco genes del IGM pueden ensayarse y utilizarse para detectar niveles de expresión que corresponden a un valor que es de "alto riesgo" (que está por encima del punto de corte) para IGM o para detectar niveles de expresión que corresponden a un valor que es de "bajo riesgo" (que se encuentra en o por debajo del punto de corte) para IGM, como se desvela en el presente documento. En algunos casos, el punto de corte del IGM puede ser 0 (cero), tal como cuando las mediciones de los niveles de expresión se normalizan a 0
45 (cero) con una desviación típica de 1. En realizaciones alternativas, el punto de corte puede estar en o aproximadamente en 0,05, en o aproximadamente en 0,10, en o aproximadamente en 0,15, en o aproximadamente en 0,20, en o aproximadamente en 0,25, en o aproximadamente en -0,05, en o aproximadamente en -0,10, en o aproximadamente en -0,15, en o aproximadamente en -0,20, en o aproximadamente en -0,25, en o aproximadamente en -0,30, en o aproximadamente en -0,35, en o aproximadamente en -0,40, en o
50 aproximadamente en -0,45, en o aproximadamente en -0,50, en o aproximadamente en -0,55, en o aproximadamente en -0,60, en o aproximadamente en -0,65, en o aproximadamente en -0,70, en o aproximadamente en -0,75, en o aproximadamente en -0,80, en o aproximadamente en -0,85, en o aproximadamente en -0,90, en o aproximadamente en -0,95, en o aproximadamente en -1,0, en o aproximadamente en -1,1, en o aproximadamente en -1,2, en o aproximadamente en -1,3, en o aproximadamente en -1,4, en o
55 aproximadamente en -1,5, en o aproximadamente en -1,6, en o aproximadamente en -1,7, en o aproximadamente en -1,8, en o aproximadamente en -1,9, en o aproximadamente en -2,0 o menos. Con respecto a el ratio H:I, su determinación puede hacerse como se describe en Ma *et al.* (2004) y Ma *et al.* (2006) como se referencia en el presente documento. Por ejemplo, puede utilizarse un valor de 0,06 para determinar si una muestra tiene una relación H:I de "alto riesgo" ($> 0,06$) o "bajo riesgo" ($\leq 0,06$).

60 De este modo, utilizando un umbral, o punto de corte, de 0 (cero) como un ejemplo no limitante para IGM con los cinco genes, los métodos divulgados proporcionan dos resultados de ensayo posibles para una muestra dada: "IGM de alto riesgo" correspondiente a un valor por encima de 0 (cero) y "IGM de bajo riesgo" correspondiente a un valor ≤ 0 . Un "IGM de alto riesgo" es indicativo de un cáncer de "alto riesgo", incluyendo el cáncer de mama, que es análogo al de un tumor de grado III como se define por las metodologías y los criterios conocidos en el campo. Un "IGM de bajo riesgo" es indicativo de un cáncer de "bajo riesgo", incluyendo el cáncer de mama, que es análogo al

de un tumor de grado I como se define por las metodologías y los criterios conocidos en el campo.

La estratificación, o clasificación, de los cánceres en dos grupos se muestra en la Figura 1, panel C y en la Figura 4, panel B, donde el nivel de riesgo identificado por un "IGM de alto riesgo" es indicativo de un aumento de la probabilidad de reaparición del cáncer, tal como la metástasis del cáncer o la reaparición distal del cáncer, incluyendo la reaparición del cáncer de mama. En muchas realizaciones, este riesgo de reaparición está presente independientemente del tratamiento con o sin tamoxifeno u otras terapias endocrinas. En las realizaciones desveladas en el presente documento, la reaparición puede ser la reaparición local del CDIS. El nivel de riesgo identificado por un "IGM de bajo riesgo" es indicativo de una menor probabilidad de reaparición del cáncer, incluyendo la reducción del riesgo de reaparición del cáncer de mama. En muchas realizaciones, el menor riesgo de reaparición está presente independientemente del tratamiento con o sin tamoxifeno. El riesgo de reaparición, o la probabilidad de reaparición, puede considerarse como el riesgo en el tiempo, tal como un período de aproximadamente 1, aproximadamente 2, aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 11, aproximadamente 12 o más años. Por tanto, la evaluación de riesgos proporcionada por el IGM puede utilizarse como un indicador de pronóstico de reaparición del cáncer y/o del resultado de supervivencia para un sujeto.

La invención incluye adicionalmente la determinación de un riesgo de reaparición en el tiempo sobre la base del valor de IGM en relación con el riesgo de reaparición determinado por las metodologías descritas en el presente documento. La Figura 9 ilustra un ejemplo no limitante de un valor de IGM de 2,1 y su indicación de un riesgo de reaparición del cáncer a los 5 años del 19 %. La figura ilustra, adicionalmente, que el riesgo de reaparición está relacionado con el valor del IGM y que la selección de 0 (cero) como umbral o punto de corte no es limitante debido a que también pueden utilizarse otros valores.

Cuando se combina con el ratio H:I, los cuatro resultados posibles del ensayo son los siguientes:

- 1) "IGM de alto riesgo" y "H:I de alto riesgo" que puede considerarse de "alto riesgo" como un "IGM de alto riesgo" por sí solo;
- 2) "IGM de alto riesgo" y "H:I de bajo riesgo" que puede considerarse como análogo a un "riesgo intermedio" de reaparición del cáncer;
- 3) "IGM de bajo riesgo" y "H:I de alto riesgo" que puede considerarse de "bajo riesgo" como un "IGM de bajo riesgo" por sí solo; y
- 4) "IGM de bajo riesgo" y "H:I de bajo riesgo" que puede considerarse de "bajo riesgo" como un "IGM de bajo riesgo" por sí solo.

La combinación de IGM y H:I por tanto identifica 3 subtipos diferentes que se ha observado que difieren en su biología tumoral y se asocian a diferentes resultados en los pacientes. Por ejemplo, un riesgo intermedio puede utilizarse para tratar al paciente con este tumor con terapia endocrina (por ejemplo, tamoxifeno como ejemplo no limitante) sobre la base de la predicción de que el paciente se beneficiará de la misma. Por el contrario, un paciente con "IGM de alto riesgo" y "H:I de alto riesgo" es poco probable que se beneficie de la monoterapia endocrina. Por tanto, la evaluación no representa un espectro continuo simple de riesgo. Esto ayuda a un médico experto, porque la evaluación identifica la biología subyacente que es útil con respecto a las opciones de tratamiento. Para hacer una selección de terapias, un médico puede determinar que cuando el paciente es de alto riesgo (es decir, alto/alto) entonces saber que este paciente es poco probable que se beneficie de la monoterapia endocrina es una información vital. Esto permite al médico considerar y/o seleccionar o aplicar una quimioterapia más agresiva o sugerir que este paciente se inscriba en un ensayo que se dirige a tumores que son resistentes a la monoterapia endocrina. La Figura 5, paneles C y D, demuestran una aplicación de estos tres grupos de riesgo en diferentes poblaciones de pacientes. Como alternativa, estas posibles combinaciones de determinaciones de IGM y H:I se utilizan como indicadores de la misma forma que el uso de IGM solo descrito anteriormente.

La capacidad del IGM para indicar el riesgo de reaparición a pesar del tratamiento con tamoxifeno también se muestra en la Figura 10, paneles A y B, donde el IGM de "alto riesgo" se utiliza como un indicador de la reaparición a pesar del tratamiento con tamoxifeno como monoterapia. Por supuesto, una combinación de determinaciones de IGM y H:I también puede utilizarse para el mismo efecto. La invención incluye adicionalmente el uso del IGM solo, de H:I sola o una combinación de IGM y H:I para predecir el grado de respuesta a un inhibidor que se dirige a los cánceres endocrino resistentes como se describe en el presente documento. La posible indicación de falta de respuesta a la terapia endocrina y de respuesta a los inhibidores divulgados puede combinarse con otro aspecto de la invención que es un método para seleccionar terapias basadas en las indicaciones de pronóstico y predictivas determinadas por los métodos divulgados. De este modo, con IGM de "alto riesgo", H:I de "alto riesgo" o una combinación de determinaciones IGM y H:I como se ha descrito anteriormente, las realizaciones de la invención incluyen métodos que incluyen adicionalmente seleccionar y, opcionalmente, tratar, el sujeto con el inhibidor para mejorar el grado de respuesta al tratamiento con tamoxifeno u otra forma de terapia endocrina. En algunos casos, el método incluye adicionalmente el tratamiento con tamoxifeno u otra forma de terapia endocrina. La invención anterior relacionada con el grado de respuesta a un inhibidor que se dirige a los cánceres endocrino resistentes también puede aplicarse en los casos de ensayo de el ratio H:I sola como se describe en el presente documento.

La invención incluye adicionalmente el ensayo de un IGM de "alto riesgo", opcionalmente con una determinación de H:I, como un indicador de la falta de respuesta a otras formas de terapia endocrina, tales como el tratamiento con un SERM, un SERD o una IA. Por supuesto, la invención también incluye la determinación de un IGM de "bajo riesgo",
 5 opcionalmente con una determinación de H:I, como un indicador del grado de respuesta al tratamiento con tamoxifeno y otros SERM, así como un SERD o un AI.

Estas posibles predicciones relativas a la terapia endocrina también pueden utilizarse en relación con los métodos para seleccionar terapias como se desvela en el presente documento. Por ejemplo, un método puede incluir no
 10 seleccionar una terapia endocrina en favor de otras terapias tales como la quimioterapia y/o la radioterapia cuando se predice la falta de respuesta. A la inversa, el método puede incluir la selección de la terapia endocrina cuando se predice la respuesta.

En realizaciones adicionales, el ensayo del IGM de "alto riesgo", el H:I de "alto riesgo" o una combinación de valores de IGM y H:I, también pueden utilizarse como un indicador del grado de respuesta relativa dentro de la terapia endocrina, tal como un mejor grado de respuesta al tratamiento con un AI con respecto a un SERM. Como ejemplo no limitante, un valor de IGM de "alto riesgo" puede utilizarse como un indicador del grado de respuesta a un IA, tal como letrozol, con respecto al tamoxifeno. Por supuesto, la invención incluye, adicionalmente, un método de selección de tratamiento con un AI basado en una predicción de este tipo.
 15
 20

Más allá de la terapia endocrina, el IGM solo, la H:I sola o una combinación de determinaciones de IGM y H:I también pueden utilizarse para indicar la falta de respuesta a la quimioterapia. Como se muestra en la Figura 11, el IGM de "alto riesgo" es predictivo de la resistencia a la quimioterapia con paclitaxel, 5-fluorouracilo, doxorubicina y ciclofosfamida como ejemplo no limitante. Por supuesto, la invención incluye adicionalmente un método de no selección de la quimioterapia como única terapia en favor de otras modalidades de tratamiento, tales como la radiación como ejemplo no limitante.
 25

Además de la terapia endocrina y la quimioterapia, la invención incluye la determinación del IGM solo, la H:I sola o una combinación de determinaciones de IGM y H:I como factor predictivo del grado de respuesta de un cáncer (sensibilidad) al tratamiento de radiación. Puede utilizarse un IGM de "alto riesgo" para predecir que un paciente con cáncer sea sensible al tratamiento con radiación, tal como después de la intervención quirúrgica. El método puede incluir adicionalmente la identificación del sujeto como que tiene probabilidades, o no tiene probabilidades, de ser sensible a la radioterapia después de la intervención quirúrgica. Por supuesto, la invención incluye, adicionalmente, un método de selección de la radioterapia sobre la base de la predicción del grado de respuesta a la misma.
 30
 35

La invención incluye adicionalmente un método de determinación del IGM solo, la H:I sola o una combinación de determinaciones de IGM y H:I como factor de pronóstico o factor predictivo de la respuesta clínica en mujeres premenopáusicas y en mujeres postmenopáusicas. La Figura 12 muestra la capacidad de un IGM de "alto riesgo" para estratificar las dos clases de mujeres sobre la base de los resultados de supervivencia. Las mujeres posmenopáusicas pueden definirse como las que tienen ≥ 50 años de edad, mientras que las mujeres premenopáusicas pueden definirse como las que tienen menos de 50 años de edad. En ambos grupos, el IGM de "alto riesgo" es un indicador del aumento de la probabilidad de reaparición del cáncer con el tiempo con respecto a un IGM de "bajo riesgo". Por supuesto, la invención incluye adicionalmente un método de no selección de terapias apropiadas para las mujeres pre y postmenopáusicas sobre la base del valor de IGM, el valor de H:I o la combinación de ambos valores determinados a partir de una muestra de la mujer.
 40
 45

Más en general, un método para determinar el tratamiento terapéutico para un paciente con cáncer puede comenzar con el ensayo de la IGM como se describe en el presente documento. El valor determinado puede utilizarse para clasificar el cáncer como correspondiente a un tumor de grado I o de grado III o identificar al sujeto como que tiene un pronóstico de reaparición probable del cáncer o que tiene respuesta o falta de respuesta a las terapias como se describe en el presente documento. El método puede entonces incluir la selección de tratamiento para un paciente con dicho tumor o dicho pronóstico o dicha respuesta o falta de respuesta. En algunos casos, el tratamiento seleccionado puede incluir la cirugía y la quimioterapia y/o la radiación debido a que el pronóstico es malo y/o se indica que no hay respuesta a otras terapias.
 50
 55

Realizaciones adicionales de la invención incluyen un método de determinación del grado tumoral o el riesgo de cáncer en un sujeto diagnosticado con cáncer benigno. El método puede incluir el ensayo de una muestra de células mamarias del sujeto para determinar los niveles de expresión de Bub1B, CENPA, NEK2, RACGAP1 y RRM2, en el que dichos niveles de expresión se correlacionan con un tumor de grado I o de grado III o un "alto riesgo" o "bajo riesgo" de cáncer. Las realizaciones de este método incluyen la determinación del valor de IGM sobre la base de los niveles de expresión y opcionalmente su uso para seleccionar el tratamiento terapéutico para el sujeto.
 60

En otras realizaciones, la invención incluye un método de determinación del grado tumoral o el riesgo de reaparición local del cáncer, en un sujeto diagnosticado con CDIS. El método puede incluir el ensayo de una muestra de células de cáncer de mama del sujeto para determinar los niveles de expresión de Bub1B, CENPA, NEK2, RACGAP1 y RRM2, en el que dichos niveles de expresión se correlacionan con un tumor de grado I o de grado III o un "alto
 65

riesgo" o "bajo riesgo" de reaparición local del cáncer.

Mientras que algunos de los anteriores se han descrito en términos de utilización de los cinco genes del IGM en combinación, la invención incluye específicamente el uso de menos de cinco, incluyendo genes individuales de entre los cinco, en la práctica de los métodos divulgados. Adicionalmente, la inclusión de otros genes con uno o más genes del IGM o el ratio H:I en lo anterior también se desvela expresamente. De forma similar, el uso de H:I con sustitución de otro índice, en su totalidad o en parte, para el gen o genes del IGM también se desvela expresamente.

Por tanto, los cinco genes del IGM pueden utilizarse individualmente con una precisión significativa o en combinación para aumentar la capacidad de correlacionar con precisión un fenotipo expresión molecular con un grado tumoral y/o un resultado del cáncer. Esta correlación es una manera de prever molecularmente la determinación de la reaparición del cáncer y/o los resultados de supervivencia como se describe en el presente documento. Hay usos adicionales de los genes correlacionados en la clasificación de células y tejidos; la determinación del diagnóstico y/o pronóstico; y la determinación y/o alteración de la terapia.

La capacidad de discriminar es conferida por la identificación de la expresión de los genes individuales como pertinente y no por la forma del ensayo utilizado para determinar el nivel real de expresión. Un ensayo puede utilizar cualquier característica de identificación de un gen individual identificado como se desvela en el presente documento siempre que el ensayo refleje, cuantitativa o cualitativamente, la expresión del gen en el "transcriptoma" (la fracción transcrita de genes en un genoma) o el "proteoma" (la fracción traducida de los genes expresados en un genoma). Las características de identificación incluyen, pero no se limitan a, secuencias de ácido nucleico únicas utilizadas para codificar (ADN) o expresar (RNA), dicho gen o epítomos específicos para, o las actividades de, una proteína codificada por dicho gen. Todo lo que se requiere es la identidad del gen o genes necesarios para discriminar entre los resultados del cáncer y una célula apropiada que contiene la muestra para su uso en un ensayo de expresión.

De forma similar, la naturaleza de la muestra que contiene células no es limitante, ya que puede utilizarse tejido fresco, tejido recién congelado y tejido fijado, tal como tejidos fijados con formol incluidos en parafina (FFIP), en los métodos divulgados.

En una realización, la invención prevé la identificación de los patrones de expresión génica mediante el análisis de la expresión génica global, o casi global, de células individuales o poblaciones de células homogéneas que se han diseccionado lejos de, o aisladas o purificadas de otro modo de, células contaminantes todo lo que sea posible mediante una simple biopsia. Debido a que la expresión de numerosos genes fluctúa entre células de diferentes pacientes, así como entre células de la muestra del mismo paciente, los niveles de expresión génica pueden determinarse en correspondencia con uno o más genes de "control" o "normalización", cuya expresión o expresiones son relativamente constantes en las células de un paciente o entre pacientes.

En otro aspecto, la invención incluye medios físicos y metodológicos para la detección de la expresión del gen o genes identificados por los modelos generados por los patrones de expresión individuales. Estos medios pueden dirigirse al ensayo de uno o más aspectos del molde o moldes de ADN subyacente a la expresión del gen o genes, del ARN utilizado como intermedio para expresar el gen o genes o del producto proteínico expresado por el gen o genes.

Una ventaja proporcionada por la invención es que las células contaminantes, no cancerosas, (tales como los linfocitos infiltrados u otras células del sistema inmunitario) no están presentes para afectar posiblemente los genes identificados o el posterior análisis de la expresión génica para identificar la reaparición del cáncer y/o los resultados de supervivencia de los pacientes. Dicha contaminación está presente cuando se utiliza una biopsia que contiene muchos tipos celulares para someter a ensayo los perfiles de expresión génica.

Si bien la presente invención se describe principalmente en el contexto del cáncer humano, tal como el cáncer de mama, puede ponerse en práctica en el contexto del cáncer de cualquier animal. Los animales preferidos para la aplicación de la presente invención son mamíferos, en particular aquellos importantes para aplicaciones agrícolas (tales como, pero no limitados a, vacas, ovejas, caballos y otros "animales de granja"), modelos animales de cáncer y animales de compañía para seres humanos (tales como, pero no limitados a, perros y gatos).

Los métodos proporcionados por la invención también pueden automatizarse en su totalidad o en parte.

Kits

Los materiales para su uso en los métodos de la presente invención son idealmente adecuados para la preparación de kits producidos de acuerdo con procedimientos bien conocidos. Por tanto, la invención proporciona kits que comprenden agentes para la detección de la expresión de los genes divulgados para la clasificación de tumores o la determinación de los resultados del cáncer. Dichos kits comprenden opcionalmente el agente con una descripción identificativa o etiqueta o instrucciones relativas a su uso en los métodos de la presente invención. Un kit de este tipo puede comprender recipientes, cada uno con uno o más de los diversos reactivos (normalmente en forma concentrada) utilizados en los métodos, incluyendo, por ejemplo, microarrays prefabricados, tampones, los

trifosfatos de nucleótidos apropiados (por ejemplo, dATP, dCTP, dGTP y dTTP; o rATP, rCTP, rGTP y UTP), transcriptasa inversa, ADN polimerasa, ARN polimerasa y uno o más complejos cebadores de la presente invención (por ejemplo, poli(T) de longitud apropiada o cebadores aleatorios unidos a un promotor reactivo con la ARN polimerasa). Normalmente también se incluirá un conjunto de instrucciones.

5 Habiendo proporcionado ahora la invención en general, la misma se entenderá más fácilmente mediante la referencia a los siguientes ejemplos que se proporcionan a modo de ilustración y que no se pretende que sean limitantes de la invención, a menos que se especifique.

10 Ejemplos

Ejemplo I: General

15 Pacientes y muestras tumorales

Se descargaron dos conjuntos de datos de microarrays de acceso público y publicados anteriormente (número de referencia GSE3494, GSE1456) de la Gene Expression Omnibus (GEO, <http://ncbi.nih.gov/geo>). GSE3494 (cohorte Uppsala) consiste en 251 pacientes procedentes de una cohorte de base poblacional tratada en el condado de Uppsala, Suecia, de 1987 a 1989, y eran heterogéneos en términos de la terapia sistémica adyuvante recibida (sin tratar o tratados con terapia endocrina y/o quimioterapia). Véase Miller LD, Smeds J, George J, *et al. Proc Natl Acad Sci USA* 102: 13550-5, 2005. Estaban disponibles los datos de los resultados clínicos (mortalidad específica por 20 cáncer de mama) para 236 pacientes con una mediana de seguimiento de 10 años. GSE1456 (cohorte Estocolmo) consiste en una serie similar de 159 pacientes con cáncer de mama tratados en el Hospital Karolinska, Estocolmo, Suecia entre 1994 y 1996. Tanto GSE3494 como GSE1456 contienen datos de expresión génica de muestras 25 tumorales congeladas analizadas en las matrices Affymetrix U133A y U133B (Affymetrix, Santa Clara, CA).

Una segunda cohorte de 239 pacientes utilizó un diseño retrospectivo de casos-cohorte (Pawitan Y, Bjohle J, Amler L, *et al. Breast Cancer Res* 7:R953-64, 2005) y se obtuvo de 683 pacientes en estadio I a estadio III, con cáncer de mama con receptor estrogénico positivo tratados en el hospital general de Massachusetts de 1991 a 1999. Los datos 30 de seguimiento clínico se obtuvieron del registro de tumores y de los registros del hospital. Los casos fueron todos los pacientes que desarrollaron metástasis a distancia durante el seguimiento; los controles se seleccionaron aleatoriamente entre los pacientes que permanecieron sin enfermedad en el último seguimiento hasta lograr una proporción de 2:1 de los controles a los casos. Además, los controles se aparearon por frecuencia con los casos con respecto a la terapia adyuvante y al tiempo de diagnóstico. Para aproximadamente el 80 % de los casos y los 35 controles, tanto los datos de resultados clínicos como los bloques de tumores fijados con formol incluidos en parafina (FFIP) se recuperaron satisfactoriamente.

La cohorte final consistía en 79 casos y 160 controles y sus características del paciente y del tumor se resumen en la Tabla 1. Este estudio fue aprobado por los comités éticos de investigación clínica locales. Esta última cohorte 40 consistía en 84 de la serie Oxford descrita anteriormente (Loi S, Haibe-Kains B, Desmedt C, *et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol* 25:1239-46, 2007). Todos los pacientes tenían cáncer de mama positivo para receptores estrogénicos y eran negativos para ganglios linfáticos y se trataron con monoterapia adyuvante con tamoxifeno. Este estudio utilizó porciones del ARN total de muestras tumorales congeladas aisladas previamente.

45

Tabla 1. Características de los pacientes y del tumor

Características	Casos (n = 79)	Controles (n = 160)
Tiempo de recaída (Años)		
Media	4,6	9,1
Intervalo	0,6-12,9	0,1-14,8
Variables apareadas Tratamiento		
quimio	11 (14 %)	9 (6 %)
quimio + endo	32 (41 %)	58 (36 %)
endo	28 (35 %)	67 (42 %)
ninguna	8 (10 %)	26 (16 %)
Año de diagnóstico		
1991-1995	40 (51 %)	74 (46 %)
1996-2000	39 (49 %)	86 (54 %)
Variables desapareadas Edad al diagnóstico (años)		
< 35	11 (14 %)	5 (3 %)
35-44	14 (18 %)	22 (14 %)
45-49	11 (14 %)	18 (11 %)
50-59	12 (15 %)	55 (34 %)
>= 60	31 (39 %)	60 (38 %)
Tamaño del tumor (cm)		

Características	Casos (n = 79)	Controles (n = 160)
< = 1	7 (9 %)	34 (21 %)
1,1-2	32 (41 %)	65 (41 %)
2,1-4	33 (42 %)	46 (29 %)
>4	7 (9 %)	15 (9 %)
Grado tumoral		
1	4 (5 %)	32 (20 %)
2	41 (52 %)	109 (68 %)
3	34 (43 %)	19 (12 %)
Estado de los ganglios linfáticos		
Neg	37 (47 %)	97 (61 %)
Pos	39 (49 %)	52 (32 %)
Desconocido	3 (4 %)	11 (7 %)
Receptor de progesterona		
Neg	19 (24 %)	20 (12 %)
Pos	60 (76 %)	140 (88 %)
Ensayos de RT-PCR en tiempo real para H/I y IGM		

Se utilizaron secuencias de cebador y sondas de HoxB13 y IL17BR, así como genes de control ESR1, PGR, CHDH, ACTB, HMBS, SDHA y UBC, como se describió anteriormente (Ma XJ, Hilsenbeck SG, Wang W, *et al J Clin Oncol* 24:4611-9, 2006). Las secuencias de cebadores y sondas para los cinco genes de grado molecular (BUB1B, CENPA, NEK2, RACGAP1 y RRM2), así como de ERBB2 (HER2) se prepararon utilizando Primer Express (ABI).

Para cada muestra FFIP, se utilizaron dos secciones de tejido de 7µm para la extracción del RNA. Se utilizó una macro-disección en bruto para enriquecer el contenido del tumor. La extracción de ARN, la transcripción inversa y la RT-PCR TaqMan utilizando el instrumento ABI 7900HT (Applied Biosystem, Inc) se realizaron como se ha descrito anteriormente (Ma *et al.*, ídem). Los números de umbral de ciclo (CT, por sus siglas en inglés) se normalizaron al CT medio de cuatro genes de referencia (ACTB, HMBS, SDHA y UBC). El uso de estos genes se apoya en los informes anteriores con respecto a estos genes y en secuencias representativas de cada uno de estos genes conocidas para el experto en la materia. Se tomaron CT normalizados para representar los niveles de expresión génica.

15 Cálculo de H/I, IGM e IGG

En general y con respecto a IGM, se prefiere que los niveles de expresión de los genes descritos se combinen para formar un único índice que sirva como un fuerte factor de pronóstico y factor predictivo del resultado o resultados clínicos. El índice es una sumatoria de los niveles de expresión de los genes utilizados y utiliza coeficientes determinados a partir de análisis de componentes principales para combinar los casos de más de un gen desvelado en un único índice. Los coeficientes se determinan mediante factores tales como la desviación típica de los niveles de expresión de cada gen en todo un conjunto de datos representativo y el valor de expresión para cada gen en cada muestra. La calidad del conjunto de datos representativo se controló sobre la base de los valores de expresión promedio para el gen o genes de referencia como se desvela en el presente documento.

25 Por ejemplo, 1

Dicho de otra manera y con respecto a IGM, los niveles de expresión normalizados para los cinco genes de los microarrays o RT-PCR se normalizaron a una media de 0 y una desviación típica de 1 en todas las muestras dentro de cada conjunto de datos y después se combinaron en un único índice por muestra a través de un análisis de componentes principales (ACP), utilizando el primer componente principal. La normalización de los datos de expresión primarios dentro de cada conjunto de datos era necesaria para representar las diferentes plataformas (microarrays y RT-PCR) y tipos de muestras (congeladas y FFIP). Como resultado, y después de los parámetros de escalado, se generó una fórmula para la sumatoria de los valores de expresión que define el índice. La precisión de los parámetros de escalado, puede entonces ensayarse basándose en las medias, los errores típicos y las desviaciones típicas (con intervalos de confianza) de los niveles de expresión de los genes en todo el conjunto de datos. Por tanto, la generación de la fórmula para el índice es dependiente del conjunto de datos, del gen de referencia y de los genes del IGM.

40 El ratio HoxB13:IL17BR se calculó como la diferencia en los niveles de expresión normalizados entre HoxB13 y IL17BR como se ha descrito anteriormente (Ma *et al.*, ídem). Las medias y las desviaciones típicas para HoxB13 y IL17BR utilizadas para la normalización de la cohorte de la Tabla 1 se obtuvieron a partir de un análisis de 190 secciones de tejido FFIP de una cohorte de base poblacional separada de los pacientes con cáncer de mama negativo para ganglios linfáticos positivo para receptor estrogénico.

45 Para el IGM, los valores de C_T en bruto obviamente anormales se eliminaron antes de hacer la media de los valores sobre los duplicados para cada gen y cada muestra. El valor de C_T en bruto promediado para cada gen después se normalizó mediante el valor de C_T promedio de los cuatro genes de referencia (ACTB, HMBS, SDHA y UBC). Los

niveles de expresión normalizados (ΔC_T) para los cinco genes se combinaron en un solo índice por muestra, que puede compararse con un valor de corte predeterminado, tal como 0, donde el IGM alto está por encima del punto de corte y el IGM bajo está por debajo del punto de corte.

- 5 El índice de grado genómico (IGG) se calculó a partir de los datos de microarrays utilizando los 128 conjuntos de sondas Affymetrix que representan 97 genes y se aumentó o redujo a escala dentro de cada conjunto de datos para que tuviera una media de -1 para los tumores de grado 1 y +1 para los tumores de grado 3 como se ha descrito anteriormente (Sotiriou *et al.*, citado anteriormente).

10 Puntos de corte y análisis estadísticos

PUNTO DE CORTE de H/I: El punto de corte de 0,06 para el ratio HoxB13:IL17BR, previamente definido para estratificar pacientes tratados con tamoxifeno adyuvante en bajo y alto riesgo de reaparición, se aplicó directamente en este estudio.

- 15 PUNTO DE CORTE de IGM: El cálculo y el punto de corte de IGM se definieron sin utilizar ningún dato de los resultados clínicos y, en su lugar, fue un punto de corte natural. El análisis inicial de IGM en la cohorte Uppsala indicó una buena discriminación de los tumores de grado 1 y de grado 3 utilizando la media (0) como punto de corte, y la agrupación basada en un modelo del IGM también indicó una distribución bimodal, con un punto de corte natural aproximadamente en 0. Este punto de corte fue apoyado adicionalmente por el análisis de la curva de característica operativa del receptor (ROC, por sus siglas en inglés).
- 20

ÍNDICE DE GRADO GENÓMICO (IGG): El IGG se dicotomizó en el punto de corte de 0, como se ha descrito anteriormente (Sotiriou *et al.*, citado anteriormente).

- 25 ANÁLISIS ESTADÍSTICOS: Se realizaron el análisis de Kaplan-Meier con ensayo de orden logarítmico y la regresión de riesgos proporcionales de Cox para evaluar la asociación de los índices de expresión génica con el resultado clínico. Se realizaron modelos de regresión multivariante de Cox para evaluar la capacidad pronóstica de los índices de expresión génica después de ajustar para factores pronósticos conocidos.
- 30

- Se comprobó el supuesto de riesgos proporcionales (PH, por sus siglas en inglés) mediante valores residuales de Schoenfeld a escala; las variables que transgredían el supuesto de PH se ajustaron en el modelo a través de la estratificación. Para representar el diseño de casos-cohortes de la cohorte de la Tabla 1, se utilizaron el análisis de Kaplan-Meier ponderado y modelos de regresión de Cox con modificaciones para manipular diseños de casos-cohortes (véase ^{19,20} como se implementó en el paquete de encuesta en R (www.r-project.org). Para someter a ensayo la interacción entre el IGM dicotomizado y el ratio H:I en modelos de regresión de Cox, se utilizó la estadística de Wald en la cohorte de la Tabla 1 y se utilizó el ensayo de razón de verosimilitud en la última cohorte.
- 35

- Las correlaciones de las variables continuas con factores categóricos se examinaron utilizando el ensayo de Wilcoxon no paramétrico de dos muestras o el ensayo de Kruskal-Wallis para los factores con más de dos niveles.
- 40

Todos los análisis estadísticos se realizaron en el entorno estadístico R. Todos los ensayos de significación fueron de dos aspectos y $p < 0,05$ se consideró significativo.

- 45 Ejemplo II: Rendimiento como pronóstico del IGM en pacientes con cáncer de mama

- La capacidad del IGM para predecir el resultado clínico en pacientes con cáncer de mama se examinó utilizando conjuntos de datos de microarrays a disposición del público. El IGM se comparó primero con el índice de grado genómico (IGG) de 97 genes descrito previamente en dos conjuntos de datos independientes. El análisis ROC indicó que el IGM y el IGG eran comparables en la discriminación de tumores de grado 1 y de grado 3 (Fig. 2). En el análisis de Kaplan-Meier, el IGM dicotomizado en el punto de corte 0 separó los pacientes en dos subgrupos con un riesgo de muerte por cáncer de mama significativamente diferente en ambos conjuntos de datos y las curvas de supervivencia y las relaciones de riesgo (RR) fueron comparables a las generados por el IGG (Fig. 2). Estos resultados demostraron, por tanto, que un índice de 5 genes podría reproducir el rendimiento como pronóstico de la firma de 97 genes mucho más compleja. Se señala que a pesar de que el IGM se desarrolló con total independencia de IGG, cuatro (BUB1B, CENPA, RACGAP1 y RRM2) de los cinco genes estaban entre la firma de 97 genes y el quinto gen, NEK2, estaba solo 2 posiciones más debajo de los 112 conjuntos de sondas asociadas al grado 3 incluidos en el IGG.
- 50
- 55

- A continuación, el IGM se examinó en un estudio TRANSBIG realizado para validar la firma de pronóstico de 76 genes Rotterdam (Desmedt C, Piette M, Loi S, *et al.*: *Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res* 13: 3207-14, 2007). Esto permitió la comparación del IGM con otra firma de pronóstico validada de forma imparcial. Con toda la cohorte, la aplicación del punto de corte 0 para el IGM dio lugar a dos grupos de pacientes con diferentes riesgos de metástasis a distancia (RR = 2,3; IC del 95 % 1,2-4,2, $p = 0,0064$), mientras que la estratificación del riesgo mediante la firma de 76 genes fue solo marginalmente significativa ($p = 0,046$). Véase la
- 60
- 65

Figura 3.

5 Además, en el subconjunto RE+ de grado 1 o 2 (n = 97), un grupo de pacientes para los que la estratificación del riesgo es más arriesgado, el IGM identificó un subgrupo de pacientes con un riesgo significativamente mayor de reaparición (RR = 3,3, IC del 95 % 1,3-8,4, p = 0,0085), mientras que la firma de 76 genes no lo hizo (RR = 1,4, p = 0,57).

10 Tomados en conjunto, en tres grandes conjuntos de datos de microarrays de un total de 608 pacientes, el IGM actuó sistemáticamente como un factor de pronóstico fuerte, comparable o superior a firmas mucho más completas.

Ejemplo III: Desarrollo y validación de un ensayo de RT-PCR para el IGM

15 Los cebadores y sondas para los 5 genes del IGM se diseñaron para el formato de ensayo de PCR en tiempo real TaqMan (RT-PCR) (Tabla 2).

Tabla 2. Secuencias de cebadores y sondas para genes del índice de grado molecular

Gen	Cebador directo	Sonda MGB Taqman	Cebador inverso
BUB1B	GCCTCAGAGCAATGGTTGTAT	ACTGTATGTGCTGTAAT	TAGTGCATCTAAATGTGTCCTAAAT
CENPA	GTGCTTGTCAACGGATGTGTAG	TCAGAACTTAATTGGG	CATCAAAGCTTACAGGTTTTCTATTCA
NEK2	CCCATGAGCCATGCCTTTC	AGTACACATGATATTTTCG	GTTGCTGAAGAACAGTAAACCCAAT
RACGAP1	GGCATCCCAACTAACAAATAAAGAG	TATAAGGGAAGATTGTCAAT	ATGACTGTAGCTTTTCTTCACACAAA
RRM2	CCTTTAACCAGCACAGCCAGTT	AAAGATGCAGCCTCA	CATTAAATCTGCCGTTGAAGCA

En comparación con las plataformas basadas en microarrays, la RT-PCR en tiempo real ofrece una mayor precisión en la cuantificación, especialmente en el análisis de muestras de ARN parcialmente degradadas de las muestras de ensayo fijadas con formol e incluidas en parafina (FFIP) (Cronin M, Pho M, Dutta D, *et al.* *Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay.* *Am J Pathol* 164:35-42, 2004), que son el tipo más común de muestra en el entorno clínico.

Para validar el ensayo del IGM basado en la RT-PCR, se realizó un estudio de casos-cohortes retrospectivo. Los casos eran pacientes que fueron tratados en el Hospital General de Massachusetts (Boston, MA) entre 1991 y 1999, pero desarrollaron metástasis a distancia durante el seguimiento, y los controles se seleccionaron aleatoriamente entre los pacientes que ingresaron en la clínica durante el mismo período y estaban libres de la enfermedad en el último seguimiento (véase la Tabla 1 anterior). Los pacientes fueron tratados con el tratamiento de referencia que no incluía ninguna terapia sistémica, terapia hormonal ni quimioterapia. Para determinar la utilidad como pronóstico independiente de la terapia del IGM, los controles con los casos con respecto a la terapia sistémica.

De manera similar a los conjuntos de datos de los microarrays analizados anteriormente, el IGM basado en RT-PCR también discriminó con precisión los tumores de grado 1 y los de grado 3 (precisión del 86 %) utilizando el mismo punto de corte 0 como se ha descrito anteriormente (Fig. 4, parte A). El análisis de Kaplan-Meier indicó que un IGM alto se asoció significativamente a un alto riesgo de metástasis a distancia, independientemente del estado ganglionar (Fig. 4, partes B-D). En un modelo de regresión de Cox multivariante ajustado para el tamaño del tumor, el grado tumoral, los ganglios linfáticos, la terapia sistémica, el IGM permaneció altamente significativo con una relación de riesgo de 4,7 (2,1-10,8) (Tabla 3).

Tabla 3. Análisis de riesgos proporcionales de Cox multivariante del IGM en toda la cohorte

Variable		Relación de riesgo	Intervalo de confianza del 95 %	p
IGM	Alto frente a bajo	4,7	2,1-10,08	0,0002
Tamaño del tumor	> 2°cm frente a < 2°cm	0,8	0,4-1,5	0,4580
Grado tumoral				0,0011
	II frente a I	1,6	0,5-5,2	0,4331
	III frente a I	5,6	1,5-20,6	0,0105
edad	> = 35 años frente a < 35 años	0,7	0,2-1,9	0,4687
Estado ganglionar	Pos. frente a Neg.	1,2	0,6-2,3	0,5581
Tratamiento				0,5733
	Quimio frente a Ninguno	0,9	0,4-2,4	0,8837
	Endo frente a, Ninguno	1,5	0,5-4,5	0,4406
	Quimio + Endo frente a Ninguno	1,0	0,3-3,5	0,9939

Por tanto, el IGM como se determinó por RT-PCR mantuvo su alta correlación con el grado tumoral y su rendimiento de pronóstico consistente en una cohorte totalmente independiente.

Ejemplo IV: Valor pronóstico complementario del IGM y HOXB13:IL17BR

Para demostrar si el ratio HOXB13:IL17BR proporciona información de pronóstico adicional al IGM y viceversa, se analizaron ambos índices en los pacientes con ganglios linfáticos negativos tratados con terapia endocrina (n = 93). Se ha demostrado que el ratio no sirve como pronóstico en pacientes con ganglios linfáticos positivos, lo cual se confirmó en esta cohorte también. En este grupo de pacientes, el IGM y el ratio H:I estaban cada uno fuertemente asociado al riesgo de metástasis a distancia (Figura 5, partes A y B).

Cuando ambos se tuvieron en cuenta juntos, el IGM fue altamente significativo en la estratificación de pacientes en grupos de bajo y alto riesgo solo cuando los tumores tenían una alta H:I, y análogamente, el ratio H:I solo fue significativa en la estratificación de pacientes con tumores que tenían un alto IGM (Figura 5E), aunque un ensayo formal para determinar la interacción entre estos dos índices no alcanzó significación estadística (p = 0,09). Por tanto, el IGM y la relación se combinaron para estratificar los pacientes en tres grupos de riesgo (bajo riesgo = bajo para ambos índices o alto solo para H:I; riesgo intermedio = alto para el IGM solamente; y alto riesgo = alto para ambos, representa el 48 %, el 24 % y el 28 % de los pacientes, respectivamente).

El análisis de Kaplan-Meier de estos tres grupos indicó que el IGM y la H:I altos juntos predijeron un muy mal resultado para el grupo de alto riesgo (relación de riesgo frente a grupo de bajo riesgo = 40,2, IC del 95 % 5,0-322,6). Esto se ilustra en la Fig. 5, parte C. Las estimaciones de Kaplan-Meier probabilidad de supervivencia sin metástasis a distancia a los 10 años fueron del 98 % (96-100 %), el 87 % (77-99 %) y el 60 % (47-78 %) para el grupo de riesgo bajo, intermedio y alto, respectivamente. Además, después de ajustar para la terapia sistémica y los factores de pronóstico convencionales (edad, tamaño y grado tumoral) en un modelo de regresión de Cox multivariante, el índice combinado se mantuvo altamente estadísticamente significativo (Tabla 4), lo que demuestra el fuerte valor pronóstico independiente de la combinación del IGM y la H:I.

Tabla 4. Modelo de riesgos proporcionales de Cox multivariante de la combinación de IGM y HOXB13:IL17BR en pacientes con ganglios negativos tratados con terapia endocrina o terapia endocrina + quimioterapia

Variables	Relación de riesgoIntervalo de confianza del 95 %p		
IGM + HOXB13:IL17BR			
Intermedio frente a Bajo	5,5	0,9-34,6	0,0007
Alto frente a Bajo	24,2	4,3-135,2	0,0003
Tamaño del tumor > 2°cm frente a <= 2°cm	1,0	0,3-2,9	0,9804
Edad >= 35 años frente a < 35 años	0,1	0,0-0,4	0,0036
Tratamiento Endo frente a Quimio + Endo	11,5	2,2-59,4	0,0034

Nota: El grado tumoral se ajustó para la estratificación.

- 5 Para fundamentar aún más el poder como pronóstico de la combinación del IGM y el ratio H:I, se examinaron estos dos índices en otra cohorte independiente de 84 pacientes RE+ con ganglios linfáticos negativos tratados de manera uniforme con terapia con tamoxifeno adyuvante (última cohorte). Después de aplicar los mismos puntos de corte a estos dos índices y el mismo algoritmo de combinación como se ha descrito anteriormente, los grupos de riesgo bajo, intermedio y alto resultantes consistieron en el 44 %, el 24 % y el 32 % de los pacientes, respectivamente, de acuerdo con sus proporciones observadas en la cohorte de la Tabla 1. De nuevo, el análisis de Kaplan-Meier indicó que el grupo de alto riesgo con tumores altos para ambos índices tuvo el peor resultado clínico (AR frente a grupo de bajo riesgo = 7,9 (2,2-28,2) (Fig. 5D) y el ensayo de razón de verosimilitud indicó una interacción estadísticamente significativa entre estos dos índices (p = 0,036).
- 10
- 15 En conjunto, en dos cohortes independientes, el IGM y el ratio H:I proporcionaron información de pronóstico adicional entre sí y la combinación de ambos índices fue particularmente eficaz en la identificación de un subconjunto de pacientes (~30 %) con muy mal pronóstico a pesar de la terapia endocrina, los que indica la necesidad de terapias adicionales para estos pacientes.

20 Ejemplo V: Correlación diferencial de HOXB13:IL17BR e IGM con expresión de RE y RP

Tanto HOXB13 como IL17BR son regulados por el receptor estrogénico. La expresión de HOXB13 se suprime mientras que la expresión de IL17BR es estimulada por los estrógenos en las estirpes celulares de cáncer de mama RE+ (Zuncaí W, S Dahiya, Provencher H, *et al. The prognostic biomarkers HOXB13, IL17BR and CHDH are regulated by estrogen in breast cancer. Clin Cancer Res*, 2007). De este modo, la expresión de los 5 genes en el IGM se sometió a ensayo para determinar su posible regulación similar por la señalización estrogénica. En el grupo de pacientes RE+ con ganglios negativos tratados con terapia endocrina analizados anteriormente, donde la H:I alta estaba fuertemente correlacionada con una menor expresión de RP, el IGM no se asoció significativamente a la expresión del mRNA de ninguno de entre el RE y el RP (Fig. 6).

30 En los tres grupos de riesgo generados por la combinación del ratio H:I y el IGM como se ha descrito anteriormente, el grupo de alto riesgo se asoció a una pobre expresión de RP. Curiosamente, el grupo de alto riesgo también estaba particularmente enriquecido en tumores que sobre-expresaban HER2 (Fig. 6). Contenía 6 de los 7 tumores con alta expresión de HER2 (utilizando 0 como punto de corte, ensayo exacto de Fisher p = 0,001). Por tanto, el grupo de alto riesgo tenía tumores tanto con la expresión de RP disminuida como con la expresión de HER2 aumentada, ambos marcadores de resistencia endocrina (Shou J, Massarweh S, Osborne CK, *et al. Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer. J Natl Cancer Inst* 96: 926-35, 2004). Esto es coherente con su mal resultado a pesar de la terapia endocrina. Estos resultados también señalan que el ratio H:I y el IGM probablemente representan vías biológicas distintas, lo que podría explicar su utilidad en la determinación de la agresividad tumoral cuando se utilizan juntos.

40 Ejemplo VI: El IGM se asocia significativamente a la RCp

45 Se utilizaron muestras tumorales de 82 pacientes RE+ tratados de forma preoperatoria con paclitaxel seguido de 5-FU, doxorubicina y ciclofosfamida (paclitaxel/FAC), una terapia que se utiliza habitualmente para el cáncer de mama, para estudiar la correlación del IGM con la sensibilidad a la quimioterapia. 7 de 82 (el 8,5 %) de los pacientes RE+ tuvieron una respuesta patológica completa (RCp) y los 7 pacientes tenían un alto IGM. Ningún tumor con un bajo IGM (~55 %) tuvo una RCp en el paciente correspondiente. Véase la Figura 11. Aproximadamente el 20 % de los tumores de alto IGM tuvieron una RCp o un enriquecimiento de 2,3 veces y un valor predictivo positivo del 100 %.

50 De este modo, el IGM alto se asocia significativamente a una respuesta patológica completa (RCp) en los tumores RE+ (p = 0,0053). Y el IGM indica qué tumores tienen probabilidad de ser sensibles o resistentes a la quimioterapia. Los resultados demuestran que el IGM y Oncotype DX tienen un rendimiento similar en la predicción de la quimio-respuesta en situaciones preoperatorias (Chang, J. *et al. "Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients." Breast Cancer Research and Treatment*, 10.1007/s10549-007-9590, 2007).

55

Bibliografía:

1. Ma XJ, Wang Z, Ryan PD, *et al* *Cancer Cell* 5:607-16, 2004
2. Ma XJ, Hilsenbeck SG, Wang W, *et al* *J Clin Oncol* 24:4611-9, 2006
- 5 3. Goetz MP, Suman VJ, Ingle JN, *et al* *Clin Cancer Res* 12:2080-7, 2006
4. Jerevall PL, Brommesson S, Strand C, *et al* *Breast Cancer Res Treat*, 2007
5. Jansen MP, Sieuwerts AM, Look MP, *et al* *J Clin Oncol* 25:662-8, 2007
6. Cianfrocca M, Goldstein LJ *Oncologist* 9:606-16, 2004
7. Sotiriou C, Wirapati P, Loi S, *et al* *J Natl Cancer Inst* 98:262-72, 2006
- 10 8. van 't Veer LJ, Dai H, van de Vijver MJ, *et al* *Nature* 415:530-6, 2002
9. Paik S, Shak S, Tang G, *et al* *N Engl J Med* 351:2817-26, 2004
10. Desmedt C, Sotiriou C *Cell Cycle* 5:2198-202, 2006
11. Loi S, Haibe-Kains B, Desmedt C, *et al* *J Clin Oncol* 25:1239-46, 2007
12. Sotiriou C, Piccart MJ *Nat Rev Cancer* 7:545-53, 2007
- 15 13. Miller LD, Smeds J, George J, *et al* *Proc Natl Acad Sci U S A* 102:13550-5, 2005
14. Pawitan Y, Bjohle J, Amler L, *et al* *Breast Cancer Res* 7:R953-64, 2005
15. Rundle AG, Vineis P, Ahsan H. *Cancer Epidemiol Biomarkers Prev* 14:1899-907, 2005
16. Ma XJ, Salunga R, Tuggle JT, *et al*. *Proc Natl Acad Sci U S A* 100:5974-9, 2003
17. Whitfield ML, Sherlock G, Saldanha AJ, *et al* *Mol Biol Cell* 13:1977-, 2000, 2002
- 20 18. Hirose, K., Kawashima, T., Iwamoto, I., *et al*. *J. Biol. Chem.* 276:5821-5828, 2001
19. Goldhirsch A, Glick JH, Gelber RD, *et al*. *Ann Oncol* 16:1569-83, 2005

Listado de secuencias

- 25 Dos secuencias representativas de ARNm de Bub1B:

ES 2 618 353 T3

1 ggtcgcttct gtagctccga gggcaggttg cggaagaaag cccaggcggg ctgtggccca
 61 gaggaaaggg ctgcagcagg acgaggacct gagccaggaa tgcaggatgg cggcggtgaa
 121 gaaggaaggg ggtgctctga gtgaagccat gtccctggag ggagatgaat gggaaactgag
 181 taaagaaaat gtacaacctt taaggcaagg gcgatcatg tccacgcttc agggagcact
 241 ggcacaagaa tctgcctgta acaatactct tcagcagcag aaacgggcat ttgaatatga
 301 aattcgattt tacactggaa atgaccctct ggatgtttgg gataggtata tcagctggac
 361 agagcagaac taccctcaag gtgggaagga gagtaatatg tcaacgttat tagaaagagc
 421 tgtagaagca ctacaaggag aaaaacgata ttatagtgat cctcgtttc tcaatctctg
 481 gcttaaatat gggcgtttat gcaatgagcc tttggatatg tacagttact tgcacaacca
 541 agggattggg gtttcacttg ctacagttcta tatctcatgg gcagaagaat atgaagctag
 601 agaaaacttt aggaaagcag atgcatatt tcaggaaggg attcaacaga aggctgaacc
 661 actagaaaga ctacagtccc agcaccgaca attccaagct cgagtgtctc ggcaaacctc
 721 gttggcactt gagaaagaag aagaggagga agtttttgag tcttctgtac cacaacgaag
 781 cacactagct gaactaaaga gcaaagggaa aaagacagca agagctcaa tcatccgtgt
 841 aggaggtgct ctcaaggctc caagccagaa cacaggactc caaatccat ttcctcaac
 901 gatgcaaaat aatagtagaa ttactgtttt tgatgaaaat gctgatgagg cttctacagc
 961 agagttgtct aagcctacag tccagccatg gatagcacc cccatgccc gggccaaaga
 1021 gaatgagctg caagcaggcc cttggaacac aggcagggtc ttggaacaca ggcctcgtgg
 1081 caatacagct tcaactgatag ctgtaccgcg tgtgcttccc agtttcactc catatgtgga
 1141 agagactgca caacagccag ttatgacacc atgtaaaatt gaacctagta taaaccacat
 1201 cctaagcacc agaaagcctg gaaaggaaga aggagatcct ctacaagggg ttcagagcca
 1261 tcagcaagcg tctgaggaga agaaagagaa gatgatgtat tgtaaggaga agatttatgc
 1321 aggagtaggg gaattctcct ttgaagaaat tgggctgaa gttttccgga agaaattaa
 1381 agagcaaagg gaagccgagc tattgaccag tgcagagaag agagcagaaa tgcagaaaca
 1441 gattgaagag atggagaaga agctaaaaga aatccaaact actcagcaag aaagaacagg
 1501 tgatcagcaa gaagagacga tgcctacaaa ggagacaact aaactgcaaa ttgcttccga
 1561 gtctcagaaa ataccaggaa tgactctatc cagttctgtt tgtcaagtaa actgttgtgc
 1621 cagagaaaact tcaacttgcg agaacatttg gcaggaacaa cctcattcta aaggcccag
 1681 tgtacctttc tccatttttg atgagtttct tctttcagaa aagaagaata aaagtcctcc
 1741 tgcagatccc ccacgagttt tagctcaacg aagaccctt gcagttctca aaacctcaga
 1801 aagcatcacc tcaaatgaag atgtgtctcc agatgtttgt gatgaattta caggaattga
 1861 acccttgagc gaggatgcca ttatcacag cttcagaaat gtaacaattt gtcctaacc
 1921 agaagacact tgtgactttg ccagagcagc tegttttgta tccactcctt tcatgagat
 1981 aatgtccttg aaggatctcc cttctgatcc tgagagactg ttaccggaag aagatctaga
 2041 tgtaaagacc tctgaggacc agcagacagc ttgtggcact atctacagtc agactctcag
 2101 catcaagaag ctgagcccaa ttattgaaga cagtcgtgaa gccacacact cctctggctt
 2161 ctctggttct tctgcctcgg ttgcaagcac ctccctccatc aaatgtcttc aaattcctga
 2221 gaaactagaa cttactaatg agacttcaga aaaccctact cagtcaccat ggtgttcaca
 2281 gtatcgcaga cagctactga agtccctacc agagttaagt gcctctgcag agttgtgtat
 2341 agaagacaga ccaatgccta agttggaaat tgagaaggaa attgaattag gtaatgagga
 2401 ttactgcatt aaacgagaat acctaatatg tgaagattac aagttattct gggtagcgcc
 2461 aagaaactct gcagaattaa cagtaataaa ggtatcttct caacctgtcc catgggactt
 2521 ttatatcaac ctcaagttaa aggaacgttt aatgaagat tttgatcatt tttgcagctg

ES 2 618 353 T3

2581 ttatcaatat caagatggct gtattgtttg gcaccaatat ataaactgct tcacccttca
 2641 ggatcttctc caacacagtg aatatattac ccatgaaata acagtgttga ttatttataa
 2701 ccttttgaca atagtggaga tgctacacaa agcagaaata gtccatggtg acttgagtcc
 2761 aagggtgtctg attctcagaa acagaatcca cgatccctat gattgtaaca agaacaatca
 2821 agctttgaag atagtggact tttcctacag tgttgacctt aggggtgcagc tggatgtttt
 2881 taccctcagc ggctttcgga ctgtacagat cctggaagga caaagatcc tggctaactg
 2941 ttcttctccc taccaggtag acctgtttgg tatagcagat ttagcacatt tactattggt
 3001 caaggaacac ctacaggtct tctgggatgg gtccttctgg aaacttagcc aaaatatattc
 3061 tgagctaaaa gatggtgaat tgtggaataa attctttgtg cggattctga atgccaatga
 3121 tgaggccaca gtgtctgttc ttggggagct tgcagcagaa atgaatgggg tttttgacac
 3181 tacattccaa agtcacctga acaaagcctt atggaaggta gggaggttaa ctagtcctgg
 3241 ggctttgctc tttcagtgag ctaggcaatc aagtctcaca gattgctgcc tcagagcaat
 3301 ggttgatttg tggaaactg aaactgtatg tgctgtaatt taatttagga cacatttaga
 3361 tgcactacca ttgctgttct actttttggg acaggatata tttgacgtca ctgatatttt
 3421 ttatacagtg atatacttac tcatggcctt gtctaacttt tgtgaagaac tattttattc
 3481 taaacagact cattacaaat ggttaccttg ttatttaacc catttgctc tacttttccc
 3541 tgtacttttc ccatttgtaa tttgtaaaat gttctcttat gatcaccatg tattttgtaa
 3601 ataataaaat agtatctggt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

1 gttagggagt cgtgtgcgtg ccttggtcgc ttctgtagct ccgagggcag gttgcggaag
 61 aaagcccagg cggctctgtgg cccagaagaa aggcctgcag caggacgagg acctgagcca
 121 ggaatgcagg atggcgggcg tgaaaaagga aggggggtgct ctgagtgaag ccatgtccct
 181 ggagggagat gaatgggaac tgagtaaaga aaatgtacaa cctttaaggc aagggcggat
 241 catgtccacg cttcagggag cactggcaca agaactctgcc tgtaacaata ctcttcagca
 301 gcagaaacgg gcatttgaat atgaaattcg attttacact ggaaatgacc ctctggatgt
 361 ttgggatagg tatatcagct ggacagagca gaactatcct caaggtggga aggagagtaa
 421 tatgtcaacg ttattagaaa gagctgtaga agcactacaa ggagaaaaac gatattatag
 481 tgatctcga tttctcaatc tctggcttaa attagggcgt ttatgcaatg agcctattgga
 541 tatgtacagt tacttgcaca accaagggat tgggtgttca cttgctcagt ctctatctc
 601 atgggcagaa gaatatgaag ctagagaaaa ctttaggaaa gcagatgcga tatttcagga
 661 agggattcaa cagaaggctg aaccactaga aagactacag tcccagcacc gacaattcca
 721 agctcgagtg tctcggcaaa ctctgttggc acttgagaaa gaagaagagg aggaagtttt
 781 tgagtcttct gtaccacaac gaagcacact agctgaacta aagagcaaag ggaaaaagac
 841 agcaagagct ccaatcatcc gtgtaggagg tgctctcaag gctccaagcc agaacagagg
 901 actccaaaat ccatttctc aacagatgca aaataatagt agaattactg tttttgatga
 961 aaatgctgat gaggcttcta cagcagagtt gtctaagcct acagtccagc catggatagc
 1021 acccccatg cccagggcca aagagaatga gctgcaagca ggcccttggga acacaggcag
 1081 gtccttgga cacaggcctc gtggcaatac agcttactg atagctgtac ccgctgtgct
 1141 tcccagtttc actccatatg tggaaagagac tgcacaacag ccagttatga caccatgtaa
 1201 aattgaacct agtataaacc acatcctaag caccagaaag cctggaaagg aagaaggaga
 1261 tcctctacaa agggttcaga gccatcagca agcatctgag gagaagaaag agaagatgat
 1321 gtattgtaag gagaagattt atgcaggagt aggggaattc tcctttgaag aaattcgggc
 1381 tgaagttttc cggaaagaaat taaaagagca aagggagcc gagctattga ccagtgcaga
 1441 gaagagagca gaaatgcaga aacagattga agagatggag aagaagctaa aagaaatcca
 1501 aactactcag caagaaagaa cagggtgatca gcaagaagag acgatgccta caaaggagac
 1561 aactaaactg caaattgctt ccgagtctca gaaaatacca ggaatgactc tatccagttc
 1621 tgtttgtcaa gtaaactggt gtgccagaga aacttactt gcggagaaca tttggcagga
 1681 acaacctcat tctaaaggtc ccagtgtaac tttctccatt tttgatgagt ttcttcttc
 1741 agaaaagaag aacaaaagtc ctctgcaga tccccacga gttttagctc aacgaagacc
 1801 ccttgcaagt ctcaaacct cagaaagcat cacctcaaat gaagatgtgt ctccagatgt

ES 2 618 353 T3

```

1861 ttgtgatgaa tttacaggaa ttgaaccctt gagcgaggat gccattatca caggcttcag
1921 aaatgtaaca atttgtccta acccagaaga cacttgtgac tttgccagag cagctcgttt
1981 tgtatccact ccttttcatg agataatgtc cttgaaggat ctcccttctg atcctgagag
2041 actgttaccg gaagaagatc tagatgtaaa gacctctgag gaccagcaga cagcttgtgg
2101 cactatctac agtcagactc tcagcatcaa gaagctgagc ccaattattg aagacagtcg
2161 tgaagccaca cactcctctg gcttctctgg ttcttctgcc tcggttgcaa gcacctcctc
2221 catcaaatgt cttcaaattc ctgagaaact agaacttact aatgagactt cagaaaaccc
2281 tactcagtca ccatgggtgt cacagtatcg cagacagcta ctgaagtccc taccagagtt
2341 aagtgcctct gcagagttgt gtatagaaga cagaccaatg cctaagttgg aaattgagaa
2401 ggaaattgaa ttaggtaatg aggattactg cattaacga gaatacctaa tatgtgaaga
2461 ttacaagtta ttctgggtgg cgccaagaaa ctctgcagaa ttaacagtaa taaaggatc
2521 ttctcaacct gtcccatggg acttttatac caacctcaag ttaaaggaac gtttaaataga
2581 agatthttgat cttttttgca gctgttatca atatcaagat ggctgtattg tttggcacca
2641 atatataaac tgcttcaccc ttcaggatct tctccaacac agtgaatata ttacctatga
2701 aataacagtg ttgattatth ataaccttht gacaatagtg gagatgctac acaaagcaga
2761 aatagtccat ggtgacttga gtccaaggtg tctgattctc agaaacagaa tccacgatcc
2821 ctatgattgt aacaagaaca atcaagctth gaagatagtg gactthtctc acagtgttga
2881 ccttagggtg cagctggatg thtttaccct cagcggctth cggactgtac agatcctgga
2941 aggacaaaag atcctggcta actgtthctc tccctaccag gttagacctg ttggtatagc
3001 agatthtagca cttthactat tgttcaagga acacctacag gtcttctggg atgggtcctt
3061 ctggaaactt agccaaaata thtctgagct aaaagatggg gaattgtgga ataaattctt
3121 tgtgctgatt ctgaatgcca atgatgaggc cacagtgtct gttcttgggg agcttgcagc
3181 aaaaatgaat ggggtthttg aactacatt ccaaagtcac ctgaacaagg ccttatggaa
3241 ggtagggag ttaactagtc ctggggctth gctctthcag tgagctaggc aatcaagtct
3301 cacagattgc tgcctcagag caatggthgt attgtggaac actgaaactg tatgtgctgt
3361 aatthaatth aggacacatt tagatgcact accgttgcct thtacttht ttggtacaggt
3421 atatthtgac gtctctgatat thtttataca gtgatatact tactcctggc ctgttctaac
3481 thttgtgaaa aactatthta thttaaacag aatcattacn aatggthacc ttgttatthta
3541 accatthgtt ctctacttht ccccgtaact thcccatthg taatthgtta aatgttctct
3601 tatgatcacc atgtatthttg taaataataa aatagtatct gthaaaaaaa aaaaaaaaaa
3661 aaaa

```

Dos secuencias representativas de ARNm de CENPA:

```

1 ccgtgaagtg ggcggagcga gcgatttgaa cgcgagcggc gcggacttct gccaaagcacc
61 ggctcatgtg aggctcgcgg cacagcgttc tctgggctcc ccagaagcca gcctthcgt
121 cccggaccgg gcagcccag caggagccgt gggaccgggc gccagcacc tctgcggcgt
181 gtcatgggccc gcgcgccggc gagccgaaag cccgaggccc cgaggaggcg cagcccagc
241 ccgaccccga cccccggccc ctcccggcgg ggcctctct taggccttc ctcccatcaa
301 cacagtgcgc ggagacaagg ttggctaaag gagatccgaa agcttcagaa gagcacacac
361 ctcttgataa ggaagctgcc cttcagccgc ctggcaagag aaatagtgt taaattcact
421 cgtgggtgtg acttcaattg gcaagcccag gccctattgg ccctacaaga ggcagcagaa
481 gcatttctag ttcatctct tgaggacgcc tatctctca ccttacctgc aggccgagtt
541 actctcttcc caaaggatgt gcaactggcc cggaggatcc ggggccttga ggagggactc
601 ggctgagctc ctgcaccag tgtttctgtc agtcttctc gctcagccag gggggatgat
661 accggggact ctccagagcc atgactagat ccaatggatt ctgcgatgct gtctggactt
721 tgctgtctct gaacagtatg tgtgtgttgc thttaaata thtctthtt ttgagaagga
781 gaagactgca tgactthtct ctgtaacaga ggtaataat gagacaatca acaccgttcc
841 aaaggcctga aaataattht cagataaaga gactccaagg ttgactthag thtgtgagtt
901 actcatgtga ctatthtgag atthtgaaaa catcagattt gctgtggtat gggagaaaa

```

5

ES 2 618 353 T3

```

961 gctatgtact tattatthta gctctttctg taatatttac atthtttacc atatgtacat
1021 ttgtacttht atthttacaca taagggaaaa aataagacca cthttgagcag ttgcctggaa
1081 ggctgggcat ttccatcata tagacctctg cccttcagag tagcctcacc attagtggca
1141 gcatcatgta actgagtgga ctgtgcttgt caacggatgt gtactthttc agaaacttaa
1201 ttggggatga atagaaaacc tgtaagctth gatgttctgg thacttctag taaattcctg
1261 tcaaaatcaa ttcagaaatt ctaacttgga gaatttaaca thttactctt gtaaatcata
1321 gaagatgtat cataacagtt cagaatthta aagtacattt tcgatgctth tatgggtatt
1381 thtgtagtht cthttgtagag agataataaa aatcaaaata thtaatgaaa a

```

```

1 cgtgaagtgg gcgagcggag cgatttgaac gcgagcggcg cggacttctg ccaagcaccg
61 gctcatgtga ggctcgcggc acagcgttct ctgggctccc cagaagccag cctthcgctc
121 ccgaccgccg cagcccgagc aggagccgtg ggaccgggcg ccagcaccct ctgcccgtg
181 tcatgggccc gcgcccggg agccgaaagc ccgaggcccc gaggaggcgc agcccagacc
241 cgaccgccgac ccccggcccc tcccggcggg gccctcctt aggcgcttcc tcccatcaac
301 acagtcggcg gagacaaggt tggctaaagg agatccgaaa gcttcagaag agcacacacc
361 tcttgataag gaagctgccc ttcagccgccc tggcagcaga agcatttcta gttcatctct
421 ttgaggacgc ctatctctct accttacctg caggccgagt tactctcttc ccaaaggatg
481 tgcaactggc ccggaggatc cggggccttg aggagggact cggctgagct cctgcacca
541 gtgtthctgt cagtctthtc tgctcagcca ggggggatga taccggggac tctccagagc
601 catgactaga tccaatggat tctgcatgct tgtctggact ttgctgtctc tgaacagtat
661 gtgtgtgttg cthtaaatat thttctthtt thtgagaagg agaagactgc atgactthcc
721 tctgtaacag aggtaatata tgagacaatc aacaccgttc caaaggcctg aaaataatth
781 tcagataaag agactccaag gttgactthta gthttgtgagt tactcatgtg actatthtgg
841 gaththtgaag acatcagatt tgctgtggta tgggagaaaa ggctatgtac thattattht
901 agctctthct gtaatathta caththttac catatgtaca thtgtactth thattthacac
961 ataagggaaa aaataagacc actthtgagca gthtgcctgga aggctgggca thtccatcat
1021 atagacctct gcccttcaga gttagcctcac cattagtggc agcatcatgt aactgagtgg
1081 actgtgcttg tcaacggatg tgtagcttht cagaaactta attggggatg aatgaaaaa
1141 ctgtaagctt tgatgttctg gthacttcta gtaaatcct gtaaaatca gtaaaatca
1201 tctaacttgg agaathtaac atthttactct tghtaatcat agaagatgta tcataacagt
1261 tcagaattht aaagtacatt ttcgatgctt thtatgggtat thttgtagth tctthttaga
1321 gagataataa aatcaaaat atthtaatgaa aa

```

Dos secuencias representativas de ARNm de NEK2:

```

1 cggggcccaa ggcaggggtg gcgggtcagt gctgctcggg ggcttctcca tccaggtccc
61 tggagthtct ggtccctgga gctccgcact tggcggcgca acctgcgtga ggcagcgcga
121 ctctggcgac tggccggcca tgcttcccg ggctgaggac tatgaagtgt tgtacaccat
181 tggcacaggc tcctacggcc gctgccagaa gatccggagg aagagtgatg gcaagatatt
241 agthttgaaa gaacttgact atggctccat gacagaagct gaaaaacaga tgcttgtthc
301 tgaagtgaat ttgcttctg aactgaaaca tccaaacatc gthcgttact atgatcggat
361 tattgaccgg accaatacaa cactgtacat tgtaatggaa tattgtgaag gaggggatct
421 ggctagtgta attacaaagg gaaccaagga aaggcaatac ttagatgaag agthttgtct
481 tcgagtgatg actcagttga ctctggccct gaaggatgac cacagacgaa gtgatggtgg
541 tcataccgta ttgcatcggg atctgaaacc agccaatgtt thcctggatg gcaagcaaaa
601 cgtcaagctt ggagactthg ggctagctag aatathaaac catgacacga gthttgcaaa
661 aacathttgt ggcacacctt attacatgct tctgaacaa atgaatcga tgtcctacaa
721 tgagaaatca gatathctgg cattgggctg cthtctgtat gagthtatgt cattaatgcc
781 tccathttaca gctthttagcc agaaagaact cgctgggaaa atcagagaag gcaaatcag

```

5

ES 2 618 353 T3

```

841 gcgaattcca taccgttact ctgatgaatt gaatgaaatt attacgagga tgttaaactt
901 aaaggattac catcgacctt ctggtgaaga aattcttgag aaccctttaa tagcagattt
961 ggttgacagac gagcaaagaa gaaatcttga gagaagaggg cgacaattag gagagccaga
1021 aaaatcgacag gattccagcc ctgtattgag tgagctgaaa ctgaaggaaa ttcagttaca
1081 ggagcgagag cgagctctca aagcaagaga agaaagattg gagcagaaag aacaggagct
1141 ttgtgttcgt gagagactag cagaggacaa actggctaga gcagaaaatc tgttgaagaa
1201 ctacagcttg ctaaaggaac ggaagtctct gtctctggca agtaatccag aacttcttaa
1261 tcttccatcc tcagtaatta agaagaaagt tcatttcagt ggggaaagta aagagaacat
1321 catgaggagt gagaattctg agagtcagct cacatctaag tccaagtgca aggacctgaa
1381 gaaaaggctt cacgctgccc agctgcgggc tcaagccctg tcagatattg agaaaaatta
1441 ccaactgaaa agcagacaga tcttgggcat gcgctagcca ggtagagaga cacagagctg
1501 tgtacaggat gtaatattac caacctttaa agactgatat tcaaatgctg tagtgttgaa
1561 tacttggttc catgagccat gcctttctgt atagtacaca tgatatttctg gaattggttt
1621 tactgttctt cagcaactat tgtacaaaa gttcacattt aatttttctt tcttcttta
1681 agaacatatt ataaaaagaa tactttcttg gttgggcttt taatcctgtg tgtgattact
1741 agtaggaaca tgagatgtga cattctaaat cttgggagaa aaaataatgt taggaaaaaa
1801 atatttatgc aggaagagta gcactcactg aatagtttta aatgactgag tggtatgctt
1861 acaattgtca tgtctagatt taaattttta gtctgagatt taaatgttt ttgagcttag
1921 aaaaccagt tagatgcaat ttggtcatta ataccatgac atcttgctta taaatattcc
1981 attgctctgt agttcaaatc tgttagcttt gtgaaaattc atcactgtga tgtttgtatt
2041 cttttttttt tttctgttta acagaatatg agctgtctgt catttaccta cttctttccc
2101 actaaataaa agaattcttc agtttccctg taaaaaaaa aaaaaaaaa aaaaaaaaa
2161 aaaaaaaaa

```

```

1 aacggggccc aaggcagggg tggcgggtca gtgctgctcg ggggcttctc catccaggtc
61 cctggagttc ctggtccctg gagctccgca cttggcggcg caacctgcgt gaggcagcgc
121 gactctggcg actggccggc catgccttcc cgggctgagg actatgaagt gttgtacacc
181 attggcacag gctcctacgg ccgctgccag aagatccgga ggaagagtga tggcaagata
241 ttagtttggg aagaacttga ctatggctcc atgacagaag ctgagaaaca ctgactgttt
301 tctgaagtga atttgcttcg tgaactgaaa catccaaaca tcgttcgtta ctatgacgg
361 attattgacc ggaccaatac aacactgtac attgtaatgg aatattgtga aggaggggat
421 ctggctagtg taattacaaa gggaaaccaag gaaaggcaat acttagatga agagtttggt
481 cttcgagtga tgactcagtt gactctggcc ctgaaggaaat gccacagacg aagtgatggt
541 ggtcataccg tattgcatcg ggatctgaaa ccagccaatg ttttctgga tggcaagcaa
601 aacgtcaagc ttggagactt tgggctagct agaatattaa accacgacac gagttttgca
661 aaaacatttg ttggcacacc ttattacatg tctcctgaac aaatgaatcg catgtcctac
721 aatgagaaat cagatatctg gtcattgggc tgcttgctgt atgagttatg tgcattaatc
781 ttttagccag aaagaactcg ctgggaaaat cagagaaggc aaattcaggc gaattccata
841 ccgttactct gatgaattga atgaaattat tacgaggatg ttaaacttaa aggattacca
901 tcgaccttct gttgaagaaa ttcttgagaa ccctttaata gcagatttgg ttgcagacga
961 gcaaagaaga aatcttgaga gaagagggcg acaattagga gagccagaaa aaaaaaaaa

```

Dos secuencias representativas de ARNm de RACGAP1:

5

```

1 ccacgcgtcc ggcggagcga agtgaagggt ggcccagggt gggccaggct gactgaaaaa
61 gatggatact atgatgctga atgtgcggaa tctgtttgag cagcttgtgc gccgggtgga
121 gattctcagt gaaggaaatg aagtccaatt tatccagttg gcgaaggact ttgaggattt
181 ccgtaaaaag tggcagagga ctgacatga gctggggaaa tacaaggatc ttttgatgaa
241 agcagagact gagcgaagtg ctctggatgt taagctgaag catgcacgta atcaggtgga

```

ES 2 618 353 T3

301 tgtagagatc aaacggagac agagagctga ggctgactgc gaaaagctgg aacgacagat
 361 tcagctgatt cgagagatgc tcatgtgtga cacatctggc agcattcaac taagcgagga
 421 gcaaaaatca gctctggcct ttctcaacag aggccaacca tccagcagca atgctgggaa
 481 caaaagacta tcaaccattg atgaatctgg ttccatttta tcagatatca gctttgacaa
 541 gactgatgaa tcaactggatt gggactcttc tttggggaag actttcaaac tgaagaagag
 601 agaaaagagg cgctctacta gccgacagtt tgttgatggt cccctggac ctgtaaagaa
 661 aactcgttcc attggctctg cagtagacca ggggaatgaa tccatagttg caaaaactac
 721 agtgactggt cccaatgatg gcggggccat cgaagctgtg tccactattg agactgtgcc
 781 atattggacc aggagccgaa ggaaaacagg tactttacaa ccttggaaaca gtgactccac
 841 cctgaacagc aggcagctgg agccaagaac tgagacagac agtgtgggca cgccacagag
 901 taatggaggg atgcgcctgc atgactttgt ttctaagacg gttattaaac ctgaatcctg
 961 tgttccatgt ggaaagcggg taaaatttgg caaattatct ctgaagtgtc gagactgtcg
 1021 tgtgggtctct catccagaat gtcgggaccg ctgtcccctt ccctgcattc ctaccctgat
 1081 aggaacacct gtcaagattg gagagggaat gctggcagac tttgtgtccc agacttctcc
 1141 aatgatcccc tccattggtg tgcattgtgt aaatgagatt gagcaaagag gtctgactga
 1201 gacaggcctg tataggatct ctggctgtga ccgcacagta aaagagctga aagagaaatt
 1261 cctcagagtg aaaactgtac ccctcctcag caaagtggat gatatccatg ctatctgtag
 1321 ccttctaaaa gactttcttc gaaacctcaa agaacctctt ctgaccttcc gccttaacag
 1381 agcctttatg gaagcagcag aaatcacaga tgaagacaac agcatagctg ccatgtacca
 1441 agctgttggg gaactgcccc aggccaacag ggacacatta gctttcctca tgattcactt
 1501 gcagagagtg gctcagagtc cacatactaa aatggatggt gccaatctgg ctaaagtctt
 1561 tggccctaca atagtggccc atgctgtgcc caatccagac ccagtgacaa tgttacagga
 1621 catcaagcgt caacccaagg tgggtgagcg cctgctttcc ttgcctctgg agtattggag
 1681 tcagttcatg atggtggagc aagagaacat tgacccccta catgtcattg aaaactcaaa
 1741 tgccttttca acaccacaga caccagatat taaagtgagt ttactgggac ctgtgaccac
 1801 tcctgaacat cagcttctca agactccttc atctagttcc ctgtcacaga gagtccggtc
 1861 caccctcacc aagaacactc ctagatttgg gagcaaaagc aagtctgcca ctaacctagg
 1921 acgacaagge aacttttttg cttctccaat gctcaagtga agtcacatct gcctgttact
 1981 tcccagcatt gactgactat aagaaaggac acatctgtac tctgctctgc agcctcctgt
 2041 actcattact acttttagca ttctccagge ttttactcaa gtttaattgt gcatgagggt
 2101 tttattaaaaa ctatatatat ctccccttcc ttctcctcaa gtcacataat atcagcactt
 2161 tgtgctgggc attggtggga gcttttagat gagacatctt tccaggggta gaagggttag
 2221 tatggaattg gttgtgattc tttttgggga agggggttat tgttcctttg gcttaagacc
 2281 aaatgctgct catagaatga tctttctcta gtttcattta gaactgattt ccgtgagaca
 2341 atgacagaaa ccctacctat ctgataagat tagcttgtct caggggtggga agtgggaggg
 2401 cagggcaaaag aaaggattag accagaggat ttaggatgcc tccttctaag aaccagaagt
 2461 tctcattccc cattatgaac tgagctataa tatggagctt tcataaaaaat gggatgcatt
 2521 gaggacagaa ctagtgatgg gagtatgctg agctttgatt tggatgatta ggtctttaat
 2581 agtgttgagt ggcacaacct tgtaaatgtg aaagtacaac tcgtatttat ctctgatgtg
 2641 ccgctggctg aactttgggt tcaattgggg tcaaagccag tttttctttt aaaattgaaat
 2701 tcattctgat gcttggcccc cataccccca acctgtcca gtggagccca acttctaag
 2761 gtcaatatat catcctttgg catcccaact aacaataaag agtaggctat aagggaagat
 2821 tgtcaatatt ttgtggtaag aaaagctaca gtcatttttt ctttgcaact tggatgctga
 2881 aatttttccc atggaacata gccacatcta gatagatgtg agctttttct tctgttaaaa
 2941 ttattcttaa tgtctgtaaa aacgattttc ttctgtagaa tgtttgactt cgtattgacc
 3001 cttatctgta aaacacctat ttgggataat atttggaaaa aaagtaaaata gctttttcaa
 3061 aatgaaaaaa aaaaaaaaa

1 gaccaggtgc gtctgccgct ggattgtgat aggaagcaga gtgttcgtgt gaaagatgga
 61 tactatgatg ctgaatgtgc ggaatctggt tgagcagctt gtgcccggg tggagattct

ES 2 618 353 T3

```

121 cagtgaagga aatgaagtcc aatztatcca gttggcgaag gactttgagg atttccgtaa
181 aaagtggcag aggactgacc atgagctggg gaaatacaag gatcttttga tgaaagcaga
241 gactgagcga agtgctctgg atgttaagct gaagcatgca cgtaatcagg tggatgtaga
301 gatcaaacgg agacagagag ctgaggctga ctgcgaaaag ctggaacgac agattcagct
361 gattcagagag atgctcatgt gtgacacatc tggcagcatt caactaagcg aggagcaaaa
421 atcagctctg gcttttctca acagaggcca accatccagc agcaatgctg ggaacaaaag
481 actatcaacc attgatgaat ctggttccat tttatcagat atcagctttg acaagactga
541 tgaatcactg gattgggact cttcttgggtg aagactttca aactgaagaa gagagaaaag
601 aggcgctcta ctagccgaca gtttgttgat ggtccccctg gacctgtaaa gaaaactcgt
661 tccatttggc tctgcagtag accaggggaa tgaatccata gttgcaaaaa ctacagtgac
721 tgttcccaat gatggcgggc ccatcgaagc tgtgtccact attgagactg tgccatattg
781 gaccaggagc cgaaggaaaa caggactttt acaaccttgg aacagtgact ccaccctgaa
841 cagcaggcag ctggagccaa gaactgagac agacagtgtg ggcacgccac agagtaatgg
901 agggatgcgc ctgcatgact ttgtttctaa gacggttatt aaactgaaat cctgtgttcc
961 atgtggaag cggataaaat ttggcaaatt atctctgaag tgtcgagact gtcgtgtggt
1021 ctctcatcca gaatgtcggg accgctgtcc ccttccctgc attcctaccc tgatagtgaa
1081 acctgtcaag attggagagg gaatgtctggc agactttgtg tcccagactt ctccaatgat
1141 cccctccatt gttgtgcatt gtgtaaataga gattgagcaa agagggtctga ctgagacagg
1201 cctgtatagg atctctggct gtgaccgcac agtaaaagag ctgaaagaga aattcctcag
1261 agtgaaaact gtacccctcc tcagcaaagt ggatgatatac catgctatct gtagccttct
1321 aaaagacttt cttcgaaacc tcaaagaacc tcttctgacc ttttcgcctt aacagagcct
1381 ttatggaagc agcagaaatc acagatgaag acaacagcat agctgccatg taccaagctg
1441 ttgggtgaact gccccaggcc aacagggaca cattagcttt cctcatgatt cacttgcaga
1501 gagtggctca gagtccacat actaaaatgg atggtgcaa tctggctaaa gtctttggcc
1561 ctacaatagt ggcccatgct gtgcccaatc cagacccagt gacaatgtta caggacatca
1621 agcgtcaacc caaggtgggt gagcgcctgc tttccttggc tctggagtat tggagtcaat
1681 tcatgatggt ggagcaagag aacattgacc ccctacatgt cattgaaaac tcaaatgcct
1741 tttcaacacc acagacacca gatattaaag tgagtttact gggacctgtg accactcctg
1801 aacatcagct tctcaagact ccttcatcta gttccctgtc acagagagtc cgttccaccc
1861 tcaccaagaa cactcctaga tttgggagca aaagcaagtc tgccactaac ctaggacgac
1921 aaggcaactt ttttgcttct ccaatgctca agtgaagtca catctgcctg ttacttcca
1981 gcattgactg actataagaa aggacacatc tgtactctgc tctgcagcct cctgtactca
2041 ttactacttt tagcattctc caggttttta ctcaagtta attgtgcata aggttttat
2101 taaaactata tatatctccc cttccttctc ctcaagtcac ataatatcag cactttgtgc
2161 tggtcattgt tgggagcttt tagatgagac atctttccag gggtagaagg gttagtatgg
2221 aattggttgt gattcttttt ggggaagggg gttattgttc ctttggctta aagccaaatg
2281 ctgctcatag aatgatcttt ctctagtttc atttagaact gatttccgtg agacaatgac
2341 agaaacccta cctatctgat aagattagct tgtctcaggg tgggaagtgg gagggcaggg
2401 caaagaaagg attagaccag aggatttagg atgcctcctt ctaagaacca gaagttctca
2461 ttccccatta tgaactgagc tataatatgg agctttcata aaaatgggat gcattgagga
2521 cagaactagt gatgggagta tgcgtagctt tgatttggat gattaggtct ttaatagtgt
2581 tgagtggcac aaccttgtaa atgtgaaagt acaactcgta tttatctctg atgtgccgct
2641 ggctgaactt tgggttcatt tggggtcaaa gccagttttt cttttaaaat tgaattcatt
2701 ctgatgcttg gccccatac cccaacctt gtccagtgga gcccaacttc taaaggtaa
2761 tatatcatcc tttggcatcc caactaacia taaagagtag gctataaggg aagattgtca
2821 atattttgtg gtaagaaaag ctacagtcac tttttctttg cactttggat gctgaaatth
2881 ttcccatgga acatagccac atctagatag atgtgagctt tttcttctgt taaaattatt
2941 cttaatgtct gtaaaaacga ttttcttctg tagaatgttt gacttcgfat tgacccttat
3001 ctgtaaaaca cctatttggg ataaaaaaaa aaaaaaaaaa aaaaa

```

Dos secuencias representativas de ARNm de RRM2:

ES 2 618 353 T3

1 cccaggecga gccaatggga agggctggag gcatggcaca gccaatggga agggccgggg
 61 caccaaagcc aatgggaagg gccgggagcg cgcggcgcgg gagatttaaa ggctgctgga
 121 gtgaggggtc gcccggtgcac cctgtcccag ccgtcctgtc ctggctgctc gctctgcttc
 181 gctgcgctc cactatgctc tcctccgtg tcccgtcgc gccatcacg gaccgcagc
 241 agctgcagct ctgcgcgctg aaggggctca gcttggtcga caaggagaac acgccgccg
 301 ccctgagcgg gaccgcgctc ctggccagca agaccgcgag gaggatcttc caggagccca
 361 cggagccgaa aactaaagca gctgcccccg gcgtggagga tgagccgctg ctgagagaaa
 421 acccccgcgg ctttgtcatc ttccccatcg agtaccatga tatctggcag atgtataaga
 481 aggcagaggc ttccttttgg accgccgagg aggttgacct ctccaaggac attcagcact
 541 gggaatccct gaaacccgag gagagatatt ttatatccca tgttctggct ttctttgcag
 601 caagcgatgg catagtaaat gaaaacttgg tggagcgatt tagccaagaa gttcagatta
 661 cagaagcccg ctgtttctat ggcttccaaa ttgccatgga aaacatacat tctgaaatgt
 721 atagtcttct tattgacact tacataaaag atcccaaaga aagggaatth ctcttcaatg
 781 ccattgaaac gatgccttgt gtcaagaaga aggcagactg ggccttgcgc tggattgggg
 841 acaaagaggc tacctatggt gaacgtgttg tagcctttgc tgcagtggaa ggcattttct
 901 tttccggttc ttttgcgtcg atattctggc tcaagaaacg aggactgatg cctggcctca
 961 cattttctaa tgaacttatt agcagagatg agggtttaca ctgtgatttt gcttgcctga
 1021 tgttcaaaca cctggtacac aaaccatcgg aggagagagt aagagaaata attatcaatg
 1081 ctgttcggat agaacaggag ttccctactg aggccttgcc tgtgaagctc attgggatga
 1141 attgcaactc aatgaagcaa tacattgagt ttgtggcaga cagacttatg ctggaactgg
 1201 gttttagcaa ggttttcaga gtagagaacc cattedgactt tatggagaat atttcaactg
 1261 aaggaaagac taacttcttt gagaagagag taggcgagta tcagaggatg ggagtgatgt
 1321 caagtccaac agagaattct tttaccttgg atgctgactt ctaaataaac tgaagatgtg
 1381 cccttacttg gctgattttt ttttccatc tcataagaaa aatcagctga agtgttacca
 1441 actagccaca ccatgaattg tccgtaatgt tcattaacag catctttaa actgtgtagc
 1501 tacctcacia ccagtcctgt ctgtttatag tgctggtagt atcacctttt gccagaaggc
 1561 ctggctggct gtgacttacc atagcagtga caatggcagt cttggcttta aagtgagggg
 1621 tgacccttta gtgagcttag cacagcggga ttaaacagtc ctttaaccag cacagccagt
 1681 taaaagatgc agcctcactg cttcaacgca gattttaatg tttacttaaa tataaacctg
 1741 gcactttaca aacaaataaa cattgttttg tactcacggc ggcgataata gcttgattta
 1801 tttggtttct acaccaaata cattctcctg accactaatg ggagccaatt cacaattcac
 1861 taagtgacta aagtaagtta aacttgtgta gactaagcat gtaattttta agttttatt
 1921 taatgaatta aaatatttgt taaccaactt taaagtcagt cctgtgtata cctagatatt
 1981 agtcagttgg tgccagatag aagacaggtt gtgtttttat cctgtggctt gtgtagtgtc
 2041 ctgggattct ctgccccctc tgagtagagt gttgtgggat aaaggaatct ctcagggcaa
 2101 ggagcttctt aagttaaatc actagaaatt taggggtgat ctgggccttc atatgtgtga
 2161 gaagccgttt cattttattt ctactgtat tttcctcaac gtctggttga tgagaaaaaa
 2221 ttcttgaaga gttttcatat gtgggagcta aggtagtatt gtaaaatttc aagtcatcct
 2281 taaacaaaat gatccacctc agatcttgcc cctgttaagt ggtgaaatca actagaggtg
 2341 gttcctacaa gttgttcatt ctagttttgt ttggtgtaag taggttgtgt gaggtaattc
 2401 atttatattt actatgtctg ttaaatacaga aattttttat tatctatggt cttctagatt
 2461 ttacctgtag ttcataaaaa aaaaaaaaaa aaaaaaaaaa

1 ccgtcctgtc ctggctgctc gctctgcttc gctgcgccgc cactatgctc tcctccgtg
 61 tcccgtcgc gccatcacg gaccgcagc agctgcagct ctgcgcgctg aaggggctca
 121 gcttggtcga caaggagaac acgccgccg ccctgagcgg gaccgcgctc ctggccagca
 181 agaccgcgag gaggatcttc caggagccca cggagccgaa aactaaagca gctgcccccg
 241 gcgtggagga tgagccgctg ctgagagaaa acccccgcgg ctttgtcatc ttccccatcg

ES 2 618 353 T3

```

301 agtaccatga tatctggcag atgtataaga aggcagaggc ttccttttgg accgccgagg
361 aggtggacct ctccaaggac attcagcact gggaatccct gaaacccgag gagagatatt
421 ttatatccca tgttctggct ttctttgcag caagcgatgg catagtaaat gaaaacttgg
481 tggagcgatt tagccaagaa gttcagatta cagaagcccg ctgtttctat ggcttccaaa
541 ttgccatgga aacatacat tctgaaatgt atagtcttct tattgacact tacataaaag
601 atcccaaaga aaggggaattt ctcttcaatg ccattgaaac gatgccttgt gtcaagaaga
661 aggcagactg ggccttgccg tggattgggg acaaagaggc tacctatggt gaacgtggtg
721 tagcctttgc tgcagtgga ggcattttct tttccgggtc ttttgcgtcg atattctggc
781 tcaagaaacg aggactgatg cctggcctca cattttctaa tgaacttatt agcagagatg
841 agggtttaca ctgtgatttt gcttgccctga tgttcaaaca cctgggtacac aaaccatcgg
901 aggagagagt aagagaaata attatcaatg ctggtcggat agaacaggag ttcctcactg
961 aggccttgcc tgtgaagctc attgggatga attgcactct aatgaagcaa tacattgagt
1021 ttgtggcaga cagacttatg ctggaactgg gttttagcaa ggttttcaga gtagagaacc
1081 catttgactt tatggagaat atttcactgg aaggaaagac taacttcttt gagaagagag
1141 taggcgagta tcagaggatg ggagtgatgt caagtccaac agagaattct tttaccttgg
1201 atgctgactt ctaaataaac tgaagatgtg cccttacttg gctgattttt tttttccatc
1261 tcataagaaa aatcagctga agtgttacca actagccaca ccatgaattg tccgtaatgt
1321 tcattaacag catctttaaa actgtgtagc tacctcacia ccagtcctgt ctgtttatag
1381 tgctggtagt atcacctttt gccagaaggc ctggctggct gtgacttacc atagcagtga
1441 caatggcagt cttggcttta aagtgagggg tgacccttta gtgagcttag cacagcggga
1501 ttaaacagtc ctttaaccag cacagccagt taaaagatgc agcctcactg cttcaacgca
1561 gattttaatg tttacttaaa tataaacctg gcactttaca aacaaataaa cattgtttgt
1621 actcacaanaa aaaaaaaaaa aaaaaaaaaa

```

REIVINDICACIONES

1. Un método que comprende:

- 5 analizar una muestra de células de cáncer de mama de un sujeto que padece cáncer de mama para determinar el nivel de expresión de Bub1B, CENPA, NEK2, RACGAP1 y RRM2 y sumar los niveles de expresión como un único índice para determinar un único índice para el sujeto;
 analizar el ratio de la expresión de HoxB13 frente a la expresión IL17BR (ratio H:I) para el sujeto en la muestra;
 comparar el índice del sujeto con un punto de corte del índice;
- 10 comparar el ratio H:I del sujeto con un punto de corte del ratio H:I; en el que el punto de corte del índice y el punto de corte del ratio H:I se determinan como los puntos de corte naturales en una distribución bimodal de índices o ratios H:I en un conjunto de datos, en el que el conjunto de datos comprende índices o ratios H:I de pacientes que (i) tuvieron reaparición del cáncer y (ii) no tuvieron reaparición del cáncer;
 clasificar el índice o el ratio H:I del sujeto como de "alto riesgo" si el índice o el ratio H:I corresponde a un valor por encima del punto de corte del índice o del punto de corte del ratio H:I o clasificar el índice o el ratio H:I del paciente como de "bajo riesgo" si el índice o el valor de H:I corresponde a un valor igual o inferior al valor de punto de corte del índice o el valor de punto de corte del ratio H:I, y
 determinar que el sujeto tiene un riesgo aumentado de reaparición del cáncer si el índice y el ratio H:I del sujeto se clasifican como de alto riesgo, o predecir que el sujeto tiene un riesgo reducido de reaparición del cáncer si el índice y el ratio H:I del sujeto se clasifican como de bajo riesgo o si el índice del sujeto se clasifica como de bajo riesgo y el ratio H:I del sujeto se clasifica como de alto riesgo, o predecir que el sujeto tiene una probabilidad intermedia de reaparición del cáncer si el índice del sujeto se clasifica como de alto riesgo y el ratio H:I del sujeto se clasifica como de bajo riesgo.
- 25 2. El método según la reivindicación 1, en el que dicho análisis comprende la preparación de ARN a partir de dicha muestra.
3. El método según la reivindicación 2, en el que dicho ARN se utiliza para la PCR (reacción en cadena de la polimerasa).
- 30 4. El método según una cualquiera de las reivindicaciones 1 a 3, en el que dicho análisis comprende el uso de un array.
5. El método según una cualquiera de las reivindicaciones 1 a 4, en el que dicha muestra se disecciona a partir de tejido retirado de dicho sujeto.
- 35 6. El método según la reivindicación 3, en el que dicha PCR es RT-PCR (transcripción inversa-PCR), opcionalmente RT-PCR en tiempo real.
- 40 7. El método según una cualquiera de las reivindicaciones 1 a 6, en el que dicha muestra es una muestra fijada con formol incluida en parafina (FFIP).
8. El método según una cualquiera de las reivindicaciones 1 a 7, en el que dicho cáncer es el carcinoma ductal *in situ* (CDIS).
- 45 9. El método según una cualquiera de las reivindicaciones 1 a 8, en el que dicha reaparición del cáncer comprende la reaparición local.
10. El método según una cualquiera de las reivindicaciones 1 a 9, que comprende adicionalmente la selección de una terapia para el sujeto sobre la base de los resultados de la determinación.
- 50 11. El método según la reivindicación 1, en el que la modelización estadística para determinar el punto de corte del índice comprende la agrupación basada en un modelo del índice para obtener una distribución bimodal del índice, y en el que el punto de corte del índice se determina como el punto de corte natural de la distribución bimodal.
- 55 12. El método según la reivindicación 11, en el que el punto de corte del índice es 0.
13. El método según la reivindicación 1, en el que la modelización estadística para determinar el punto de corte del ratio H:I comprende el análisis de regresión logística del conjunto de datos.
- 60 14. El método según la reivindicación 13, en el que el punto de corte del ratio H:I es 0,06.

Figura 1

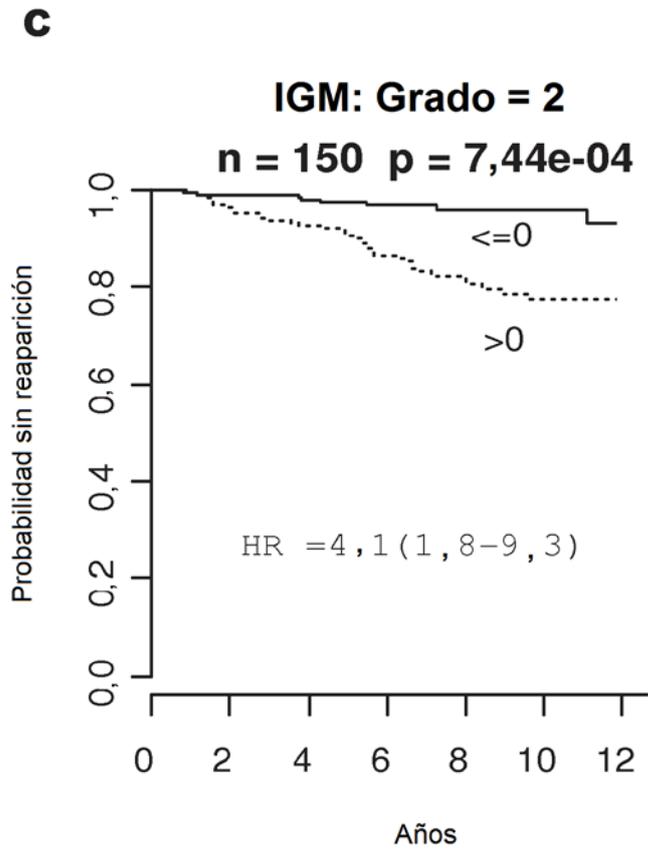
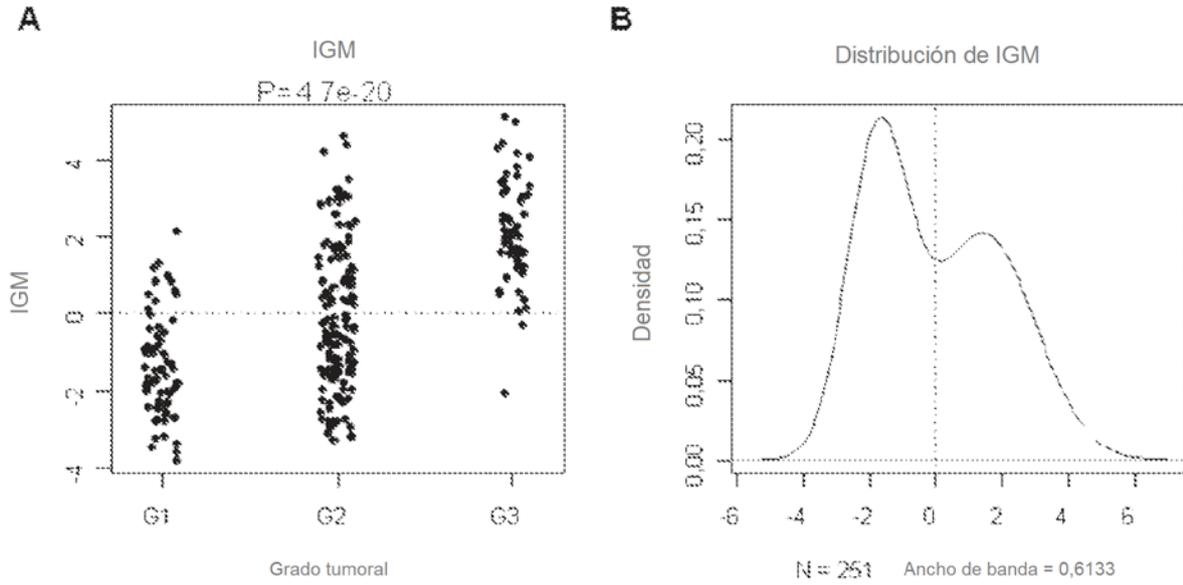


Figura 2

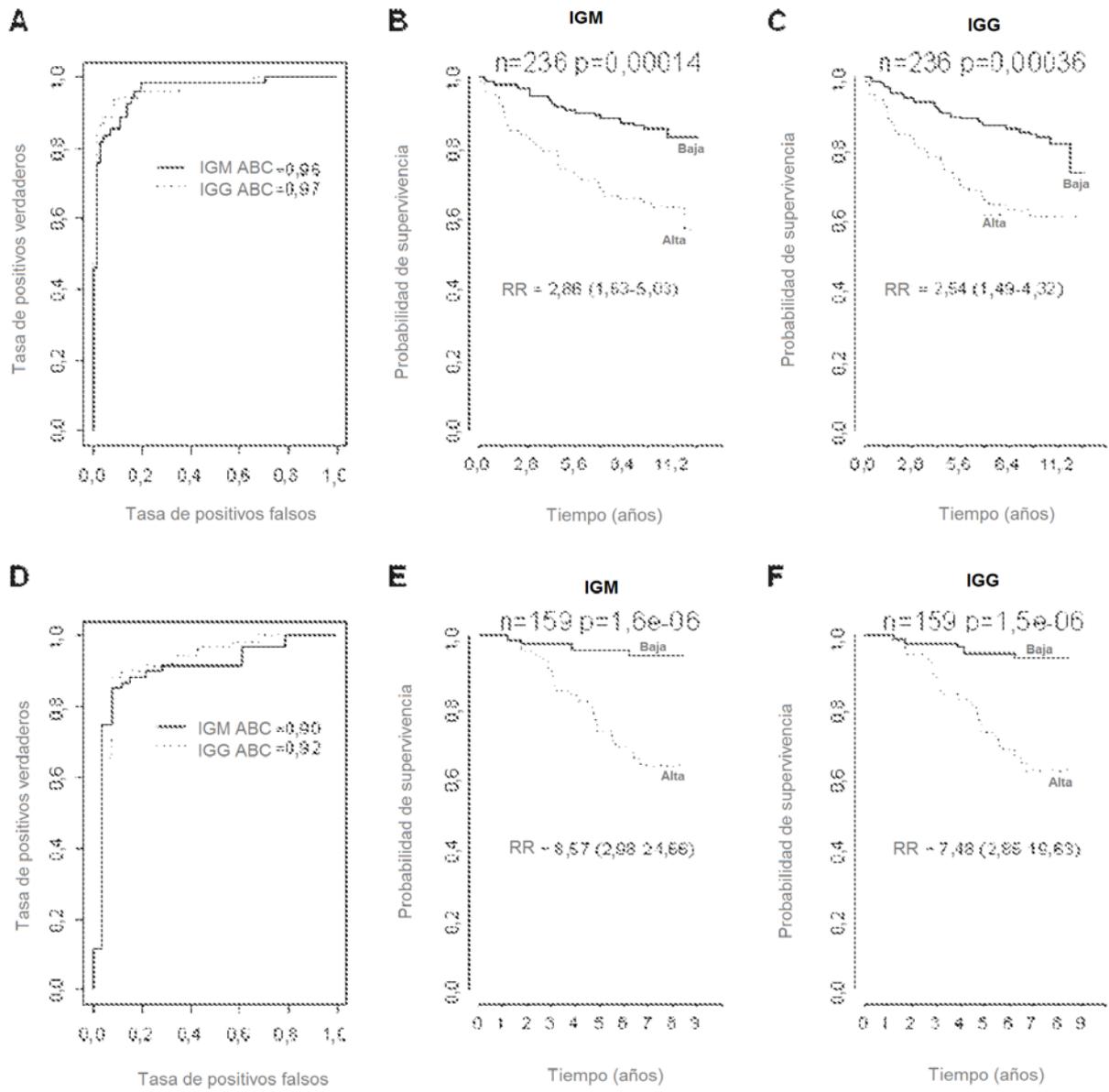


Figura 3

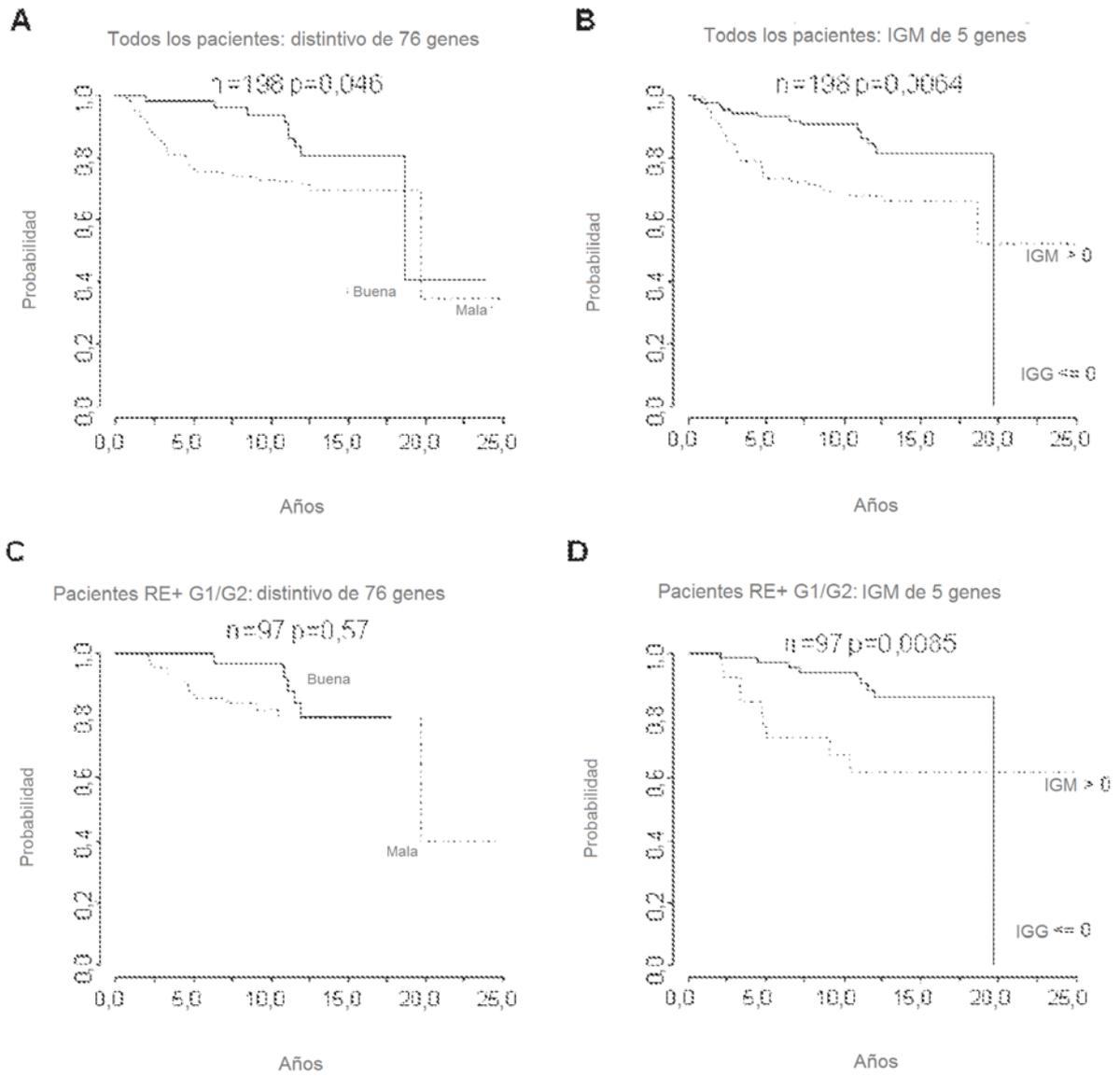


Figura 4

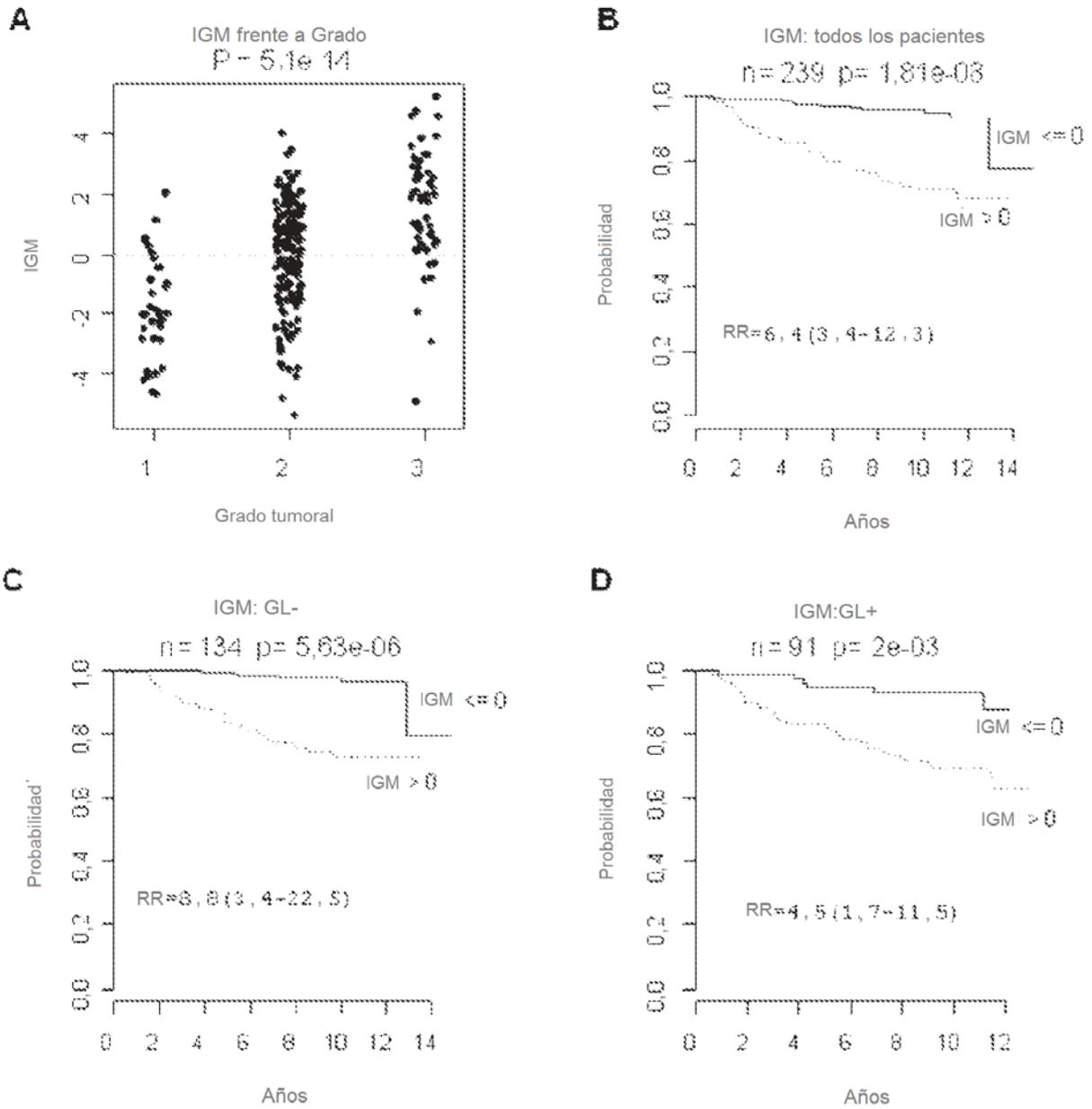


Figura 5

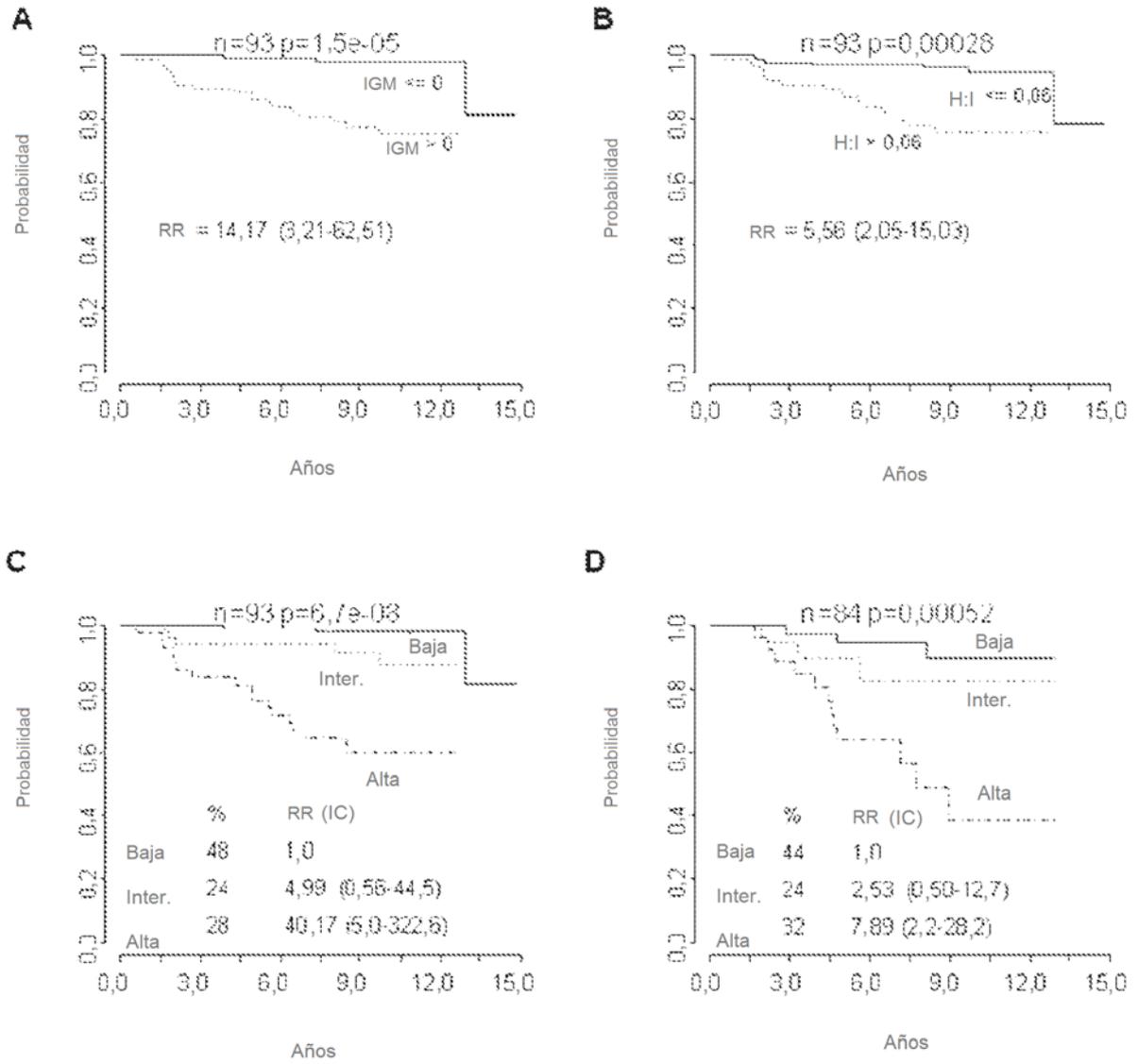


Figura 5E

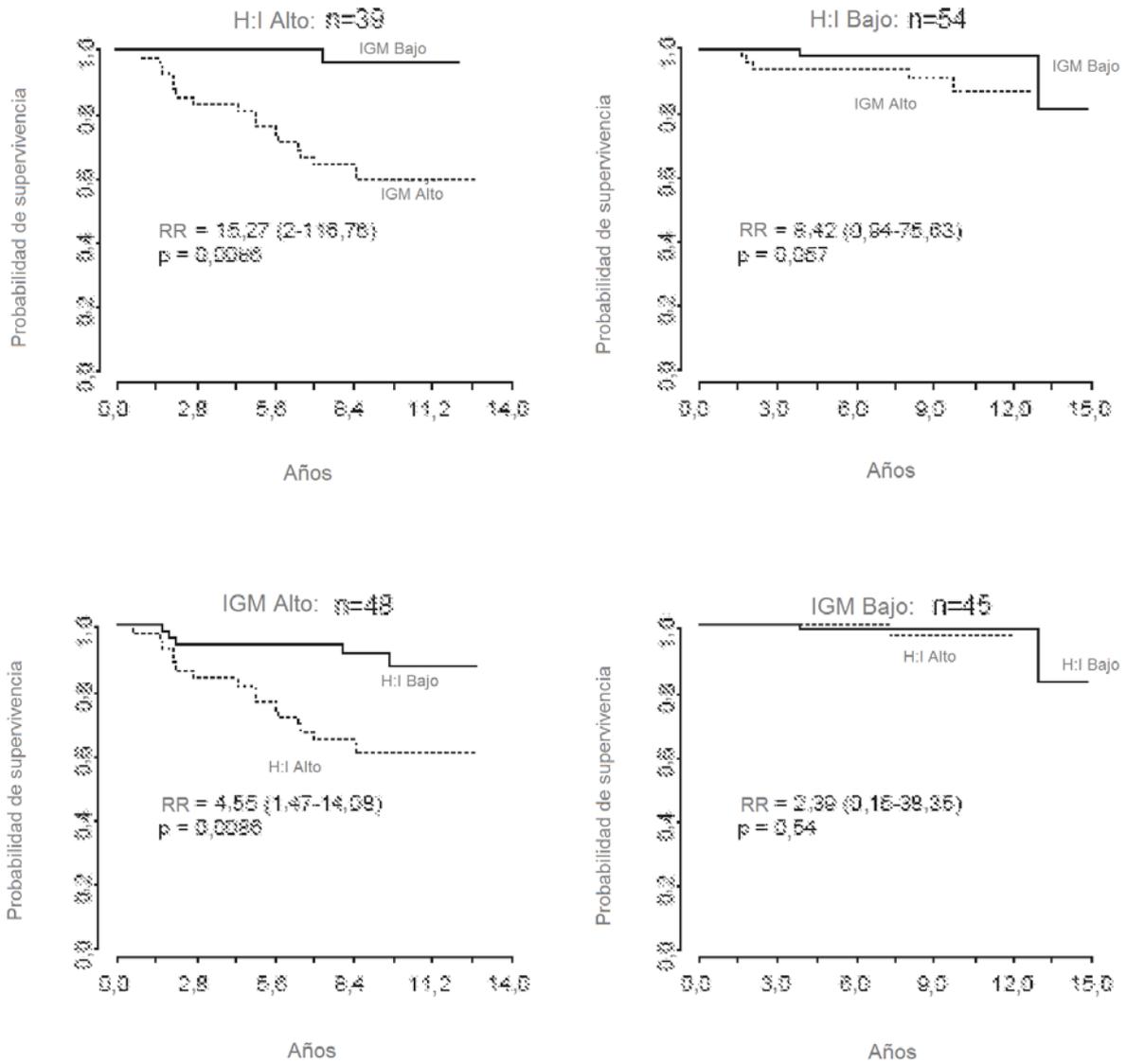


Figura 6

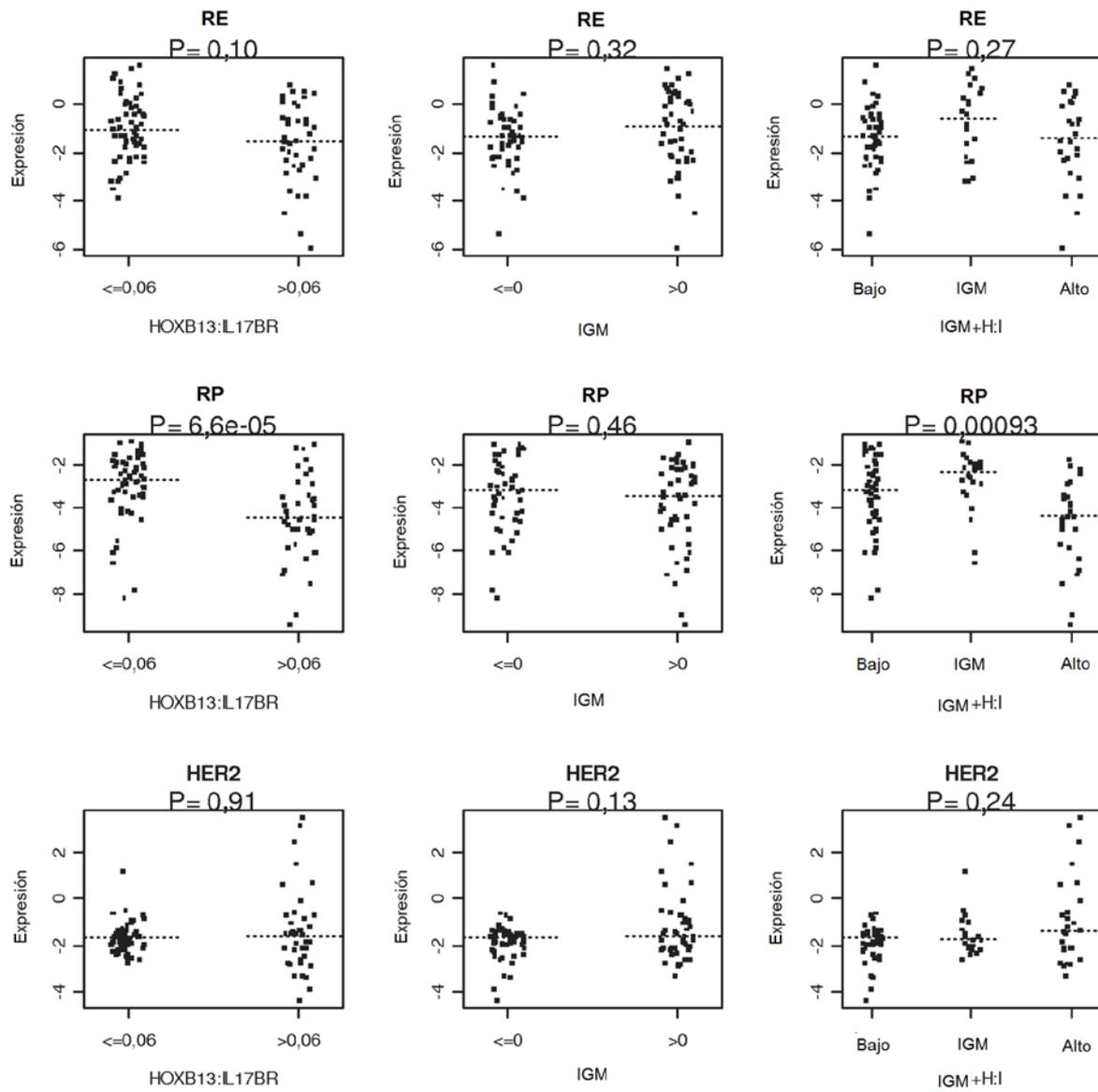


Figura 7

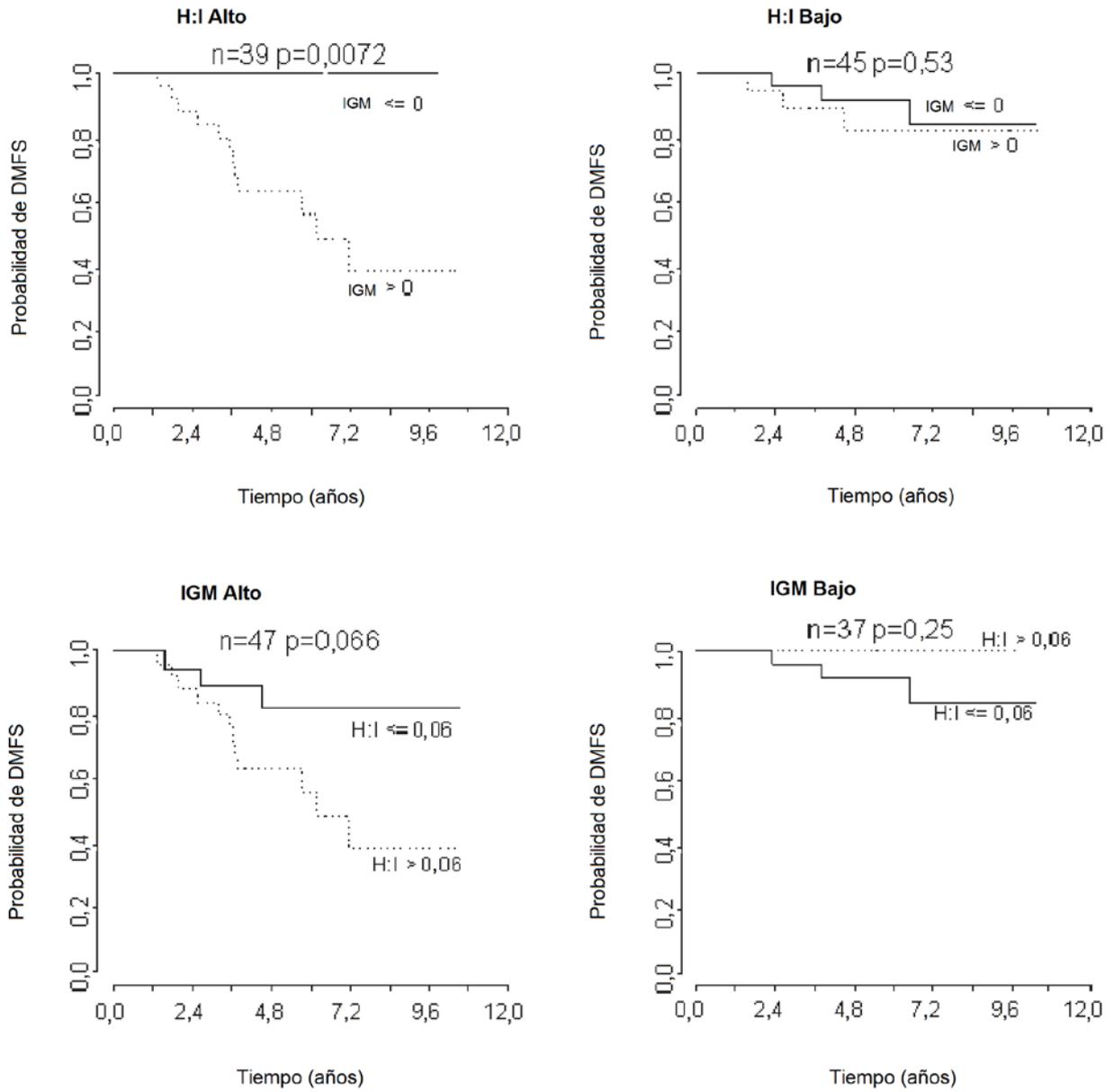


Figura 8

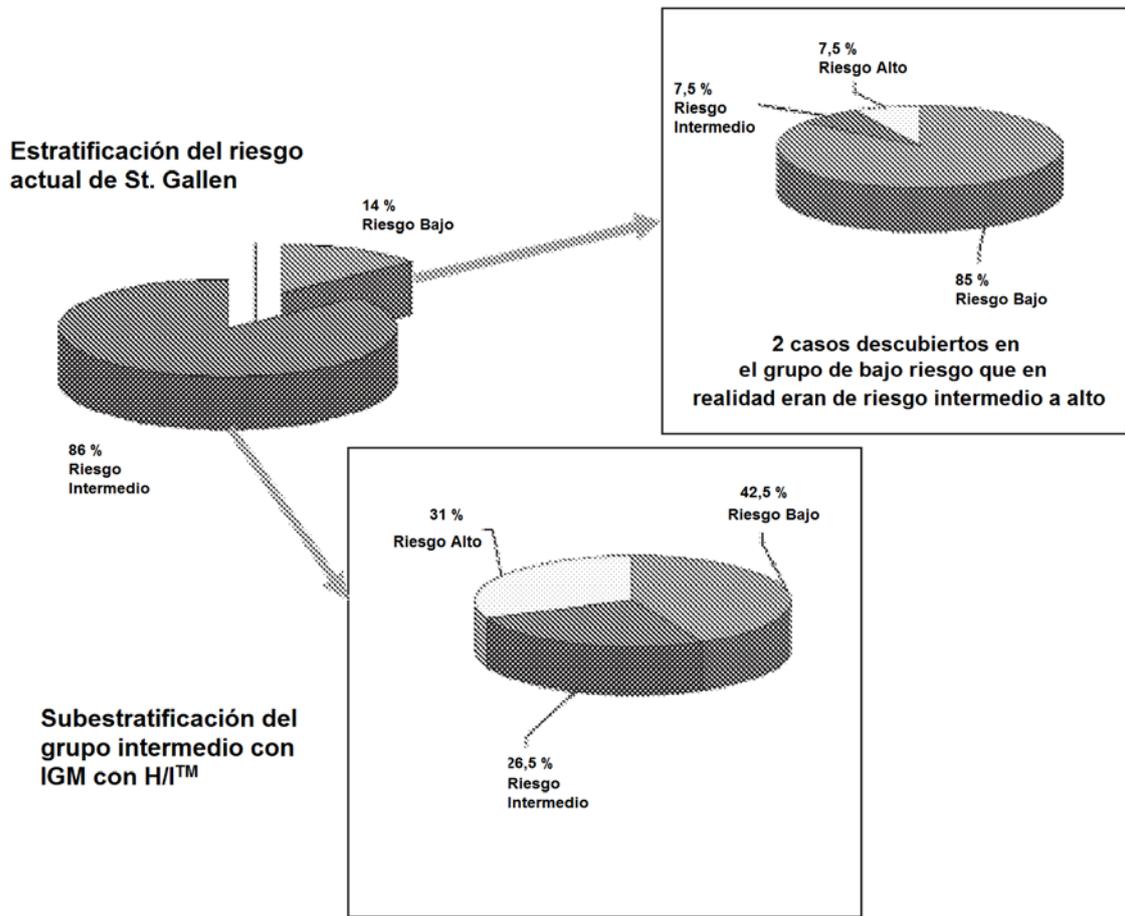


Figura 9

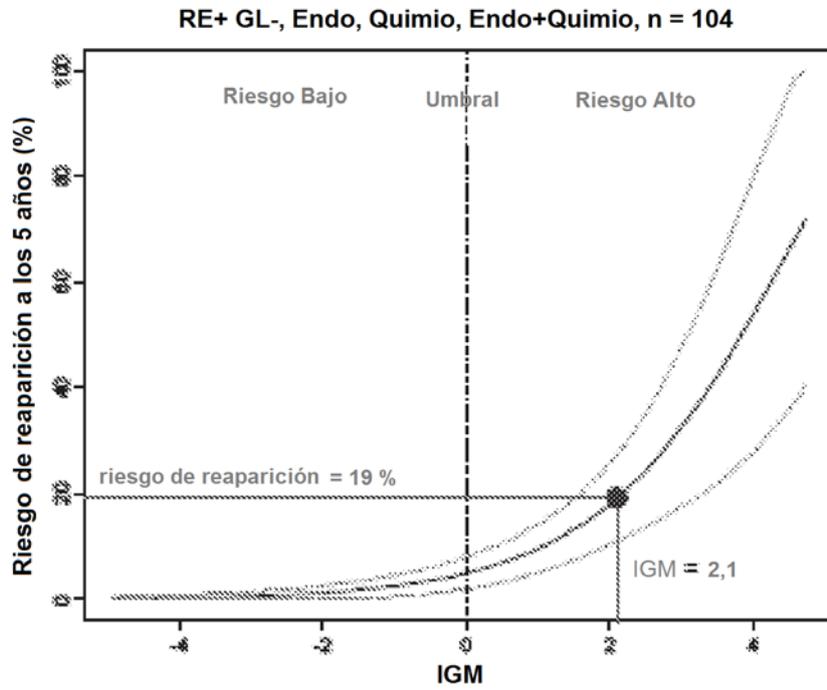


Figura 10

